THE UNIVERSITY OF CHICAGO

ASTRONOMY AND COSMOLOGY WITH GRAVITATIONAL WAVES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ASTRONOMY & ASTROPHYSICS

BY
MAYA FISHBACH

CHICAGO, ILLINOIS
AUGUST 2020

To my parents **לאמא ואבא**

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to thank the following people, without which I would not have been able to complete my dissertation:

Thank you to my advisor, Daniel Holz, for his guidance and support over the course of this PhD, and for teaching me to ask interesting questions. Thank you to my thesis committee, Daniel Holz, Steve Meyer, Leslie Rogers and Bob Wald. Thank you to my frequent collaborator and mentor Will Farr for his patient and careful explanations, and for many statistics lessons.

Thank you to the University of Chicago LIGO group, Hsin-Yu Chen, Zoheyr Doctor, Reed Essick, Jose Maria Ezquiaga, Amanda Farah, Ben Farr, and Phil Landry, for many discussions, scientific and non-scientific, and for their friendship. Thank you to the members of the University of Chicago Astronomy & Astrophysics Department and the Kavli Institute for Cosmological Physics, especially the graduate students, postdocs, administrators, staff and faculty who work to improve our community.

Thank you to my parents Ayelet and Alon, who instilled in me my love of learning, and always encouraged me to pursue my dreams (especially if those dreams involved attaining a PhD). Thank you to my siblings Shira and Tomer for always making me laugh. Thank you to my dog Dobby and hamster Mitski for being part of our loving and caring home.

Finally, I would like to acknowledge that the last few months of my PhD were completed during the COVID-19 pandemic. Thank you to all of the essential workers on the frontline of this pandemic. Thank you to everyone who is fighting to save lives and to make the world a better, more just place.

ABSTRACT

Starting with the first gravitational-wave detection in September 2015, the LIGO and Virgo gravitational-wave detectors are revealing a new astrophysical population of merging binary black holes and neutron stars. This dissertation focuses on the astrophysical and cosmological lessons enabled by the rapidly growing catalog of gravitational-wave detections. In the first part of the dissertation, we study the properties of the binary black hole population, including the shape of the black hole mass function, the distribution of spins, and the merger rate and its evolution in cosmic time. We show that measuring the spins of the black holes in a merging system can reveal whether they themselves formed from previous mergers. Analyzing the masses of the black holes detected by LIGO, we show the first evidence for missing black holes in the mass range ~ 50 – $100\ M_{\odot}$, and discuss the implications for stellar evolution and supernova theory. We also analyze the pairing of component black holes in a binary, and find that black holes tend to merge with similar mass partners rather than pairing randomly. Additionally, we analyze the binary black hole merger rate as a function of redshift and the implications for the formation rate of the progenitor stars and the time delays between formation and merger. The second part of the dissertation focuses on gravitational-wave cosmology. We explore the potential of gravitational-wave signals to measure the Hubble constant, and perform the first such statistical standard siren measurement of the Hubble constant using gravitational-wave data together with a galaxy catalog.

CHAPTER 1

INTRODUCTION

A century after Einstein's prediction, we are finally listening to the universe in gravitational waves. These waves are tiny ripples in spacetime that can be detected by the Laser Interferometer Gravitational-Wave Observatory (LIGO), arguably the most sensitive instrument ever built. When advanced LIGO first turned on, it heard two black holes, each 30 times the mass of the Sun and located over a billion lightyears away, colliding into each other at roughly half the speed of light. This event was dubbed GW150914. In the last four years, we have detected and published nine additional collisions of black holes and one neutron star merger. This dissertation focuses on the astrophysical and cosmological lessons enabled by the rapidly-growing catalog of gravitational-wave events, including probing the formation of black holes and the expansion of the universe.

Gravitational waves from merging black holes and neutron stars carry information about these sources' properties, including their masses, spins, distances and positions on the sky. These properties are encoded in the gravitational-wave signal according to the theory of general relativity. By analyzing data from the LIGO and Virgo detectors, we can extract information about the properties of individual detected systems. In this work, rather than focusing on individually detected systems, we reconstruct the *population properties* of the black holes detected by LIGO/Virgo. Population properties are properties that are shared across the entire sample; for example, the maximum black hole mass or the merger rate as a function of redshift. By studying the underlying distribution of masses, spins and redshifts of the merging black hole population, we start to understand where black holes are created and how they merge.

In Chapters 2–5, we discover a number of significant features in the black hole population and discuss their astrophysical implications. We address questions such as: Did the observed black holes form directly from collapsed stars, or did they form from previous mergers of

smaller black holes (Chapter 2)? Is there a maximum black hole mass for black holes formed from stellar collapse (Chapter 3)? Does the black hole merger rate evolve with redshift (Chapter 4)? Do black holes pair off randomly, or do they preferentially partner with similar mass companions (Chapter 5)?

Before the first gravitational-wave direct detection in September 2015, there had been no observation of stellar-mass black holes with masses greater than $\sim 20\ M_\odot$. This first gravitational-wave signal, produced by a merger of two 30 solar mass black holes, prompted both the question "How does nature make such big ($\sim 30\ M_\odot$) black holes?" and a possible solution: one way of making bigger black holes is by merging smaller ones. In Chapter 2, we develop an analysis to test the hypothesis that the component black holes in a binary are themselves made out of smaller black holes merging. We show that black holes that have previously undergone one or more mergers tend to spin at around 70% of the maximum possible black hole spin. We find that once we have a few hundred observations of merging black holes, this feature in the spin distribution alone is enough to deduce whether some of the black holes are formed from mergers of smaller black holes.

While the black holes discovered by LIGO/Virgo are bigger than any of the stellar-mass black holes previously detected, this is unsurprising given that LIGO's sensitivity increases dramatically as a function of black hole mass (up to $\sim 100\ M_\odot$). In fact, Chapter 3 points out that LIGO's first few black hole detections are surprisingly *small*. Why have we detected mergers involving $30\ M_\odot$ black holes, but none involving $60\ M_\odot$ black holes, when LIGO is roughly 5 times more sensitive to the latter? In Chapter 3, we find that these missing big black holes provide evidence for a cutoff at $\sim 45\ M_\odot$ in the black hole mass spectrum. Stars with cores between $\sim$ 50–135 $M_\odot$ are thought to undergo pair instability at the end of their evolution and explode violently in pair-instability/ pulsational pair-instability supernovae. Such supernova explosions are associated with considerable mass loss, so that the star either leaves behind no remnant, or leaves behind a black hole smaller than $\sim 50\ M_\odot$. In addition

to revealing the physics of stellar deaths, constraining the upper end of the black hole mass distribution puts limits on the occurrence rate of second-generation mergers that may occupy the pair-instability mass gap. This has implications for whether black holes are formed in isolation (where second-generation mergers would be impossible) or in dense clusters.

Another way to learn about the formation history of black hole mergers is by examining when and where in the universe they merged, i.e. their redshifts. In Chapter 4, we perform the first measurement of the black hole merger rate as a function of redshift. The redshift distribution of black hole mergers depends on a combination of the progenitor formation rate (e.g. the star formation rate) and the delay times between formation and merger. We find that the time delay distribution can be constrained with a few hundred LIGO/Virgo detections (to be expected in the next few years).

In Chapter 5, we point out another feature in the black hole mass distribution: the first ten detected black hole binaries are all made up of equal or near-equal mass components. Analyzing the first ten LIGO-Virgo binary black holes, we find that merging black holes are unlikely to be randomly drawn from an underlying distribution, but rather prefer partners of a comparable mass. We discuss what the inferred pairing function implies for binary physics.

The population analyses presented in Chapters 2-5 rely on a proper treatment of statistical uncertainties and selection effects in order to avoid biased inference. In Chapter 6, we discuss the statistical tools of population analysis in detail, including a review of hierarchical Bayesian analysis and posterior predictive checks. We address the issue of disentangling true population outliers from statistical fluctuations, focusing on the measurement of the maximum black hole mass. We show that statistical fluctuations often cause us to overestimate the true mass of a binary black hole detected in gravitational waves, but that analyzing all events jointly in a population analysis enables us to more tightly measure the masses of individual events and assess whether there are outliers or sub-populations present.

In Chapter 7, we use gravitational-wave signals as cosmological probes. The gravitational-

wave signal from a compact binary merger directly encodes the distance to the source, lending them the name "standard sirens" in analogy with "standard candles." In order to measure the size of universe as a function of time (i.e. cosmography), one must infer both a distances and a redshift for a collection of sources. There are several different ways of combining the gravitational-wave distance measurement to a source with a redshift measurement, using (a) an electromagnetic counterpart, (b) a catalog of potential host galaxies, and (c) isolating the effects of cosmology on the population distributions of gravitational-wave sources. Chapter 7 pursues the second method, and performs the first analysis which combines gravitational-wave data and a galaxy catalog to infer the Hubble constant.

CHAPTER 2

ARE LIGO'S BLACK HOLES MADE FROM SMALLER BLACK HOLES?

The first gravitational waves directly detected were produced by a pair of $\sim 30\ M_\odot$ black holes, which collided and produced a black hole of roughly $\sim 60\ M_\odot$. This naturally led us to wonder whether these $30\ M_\odot$ black holes are themselves the merger products of smaller black holes. When two black holes merge, the resulting black hole is bigger – roughly the sum of the component masses, with some small fraction of the total mass radiated away in gravitational waves. Notably, the resulting black hole is also spinning. Regardless of how fast the initial black holes are spinning, as they inspiral together on their collision course, the system carries orbital angular momentum that gets transferred to the final black hole upon merger. There are several formulas in the literature that calculate the spin of the final black hole based on the masses and spins of the initial black holes. These formulas are based on a combination of analytic post-Newtonian considerations as well as fits to numerical relativity; however, the basic idea is that the spin of the final black hole is a sum of the spins of the initial black holes and the orbital angular momentum. As long as the initial black holes are of comparable mass, the orbital angular momentum dominates the contribution to the final spin, leading to a "universal distribution" of spins in black holes born as merger products. Black holes that are made from smaller black holes tend to spin at around 70% of the maximum possible black hole spin.

In this chapter, we point out that a universal spin distribution exists for hierarchically-formed black holes (i.e. black holes produced by mergers of smaller black holes). We then propose to use this unique spin distribution as a signature of hierarchical formation. Although we can very rarely measure the spins of individual black holes accurately enough to definitively tell whether an individual black hole is spinning at exactly 70% (implying that it is a strong candidate for hierarchical formation), we can search for evidence of hierarchical

formation in the *population* of black holes. We find that with a few hundred binary black holes detected in gravitational waves, we can measure the distribution of spins well enough to infer the fraction of merging black holes that formed in mergers of smaller black holes.

This chapter was written in collaboration with Daniel Holz and Ben Farr and published in the Astrophysical Journal Letters [106]. The authors were supported by NSF CAREER grant PHY-1151836. They were also supported in part by the Kavli Institute for Cosmological Physics at the University of Chicago through NSF grant PHY-1125897 and an endowment from the Kavli Foundation.

2.1 Abstract

One proposed formation channel for stellar mass black holes (BHs) is through hierarchical mergers of smaller BHs. Repeated mergers between comparable mass BHs leave an imprint on the spin of the resulting BH, since the final BH spin is largely determined by the orbital angular momentum of the binary. We find that for stellar mass BHs forming hierarchically the distribution of spin magnitudes is universal, with a peak at $a \sim 0.7$ and little support below $a \sim 0.5$. We show that the spin distribution is robust against changes to the mass ratio of the merging binaries, the initial spin distribution of the first generation of BHs, and the number of merger generations. While we assume an isotropic distribution of initial spin directions, spins that are preferentially aligned or antialigned do not qualitatively change our results. We also consider a "cluster catastrophe" model for BH formation in which we allow for mergers of arbitrary mass ratios and show that this scenario predicts a unique spin distribution that is similar to the universal distribution derived for major majors. We explore the ability of spin measurements from ground-based gravitational-wave (GW) detectors to constrain hierarchical merger scenarios. We apply a hierarchical Bayesian mixture model to mock GW data and argue that the fraction of BHs that formed through hierarchical mergers will be constrained with $\mathcal{O}(100)$ LIGO binary black hole detections, while with

$\mathcal{O}(10)$ detections we could falsify a model in which all component BHs form hierarchically.

2.2 Introduction

LIGO's first detections of gravitational waves (GWs) from binary black hole (BBH) systems allow us to probe the formation histories of stellar mass binary black holes (BHs; 2). Various formation channels have been proposed for the component black holes in these binaries [39, 212, 79, 48, 73, 29, 141], and these can be broadly separated into two classes: isolated binary evolution and dynamical binary formation channels that involve first-generation BHs (e.g., resulting from stellar collapse) and dynamical formation channels that involve BHs built up from the mergers of earlier generations of BHs. In this work, we consider the latter group: BH formation through hierarchical mergers wherein a BH in a BH binary is produced by a merger of two smaller BHs from a previous generation, and the previous generation's BHs may themselves be merger products of an even earlier generation. Hierarchical mergers may occur in high-density environments where some fraction of merger products do not escape despite receiving recoil kicks [188] and may therefore undergo another merger. The hierarchical merger scenario has been proposed in the context of dynamical formation in nuclear star clusters [29], young stellar clusters [183], AGN disks [187], as well as in the formation of primordial BHs [73].

In anticipation of future LIGO BBH detections, we describe a method to determine whether or not the observed BHs formed hierarchically. In particular, the hierarchical formation channel can be probed by analyzing the distribution of observed spin magnitudes of the component BHs.

Each BH in a binary has a mass m_i $(i = 1, 2)$ and spin

$$\mathbf{S_i} = a_i \frac{G m_i^2}{c} \hat{\mathbf{S}}_\mathbf{i}, \tag{2.1}$$

where a_i is the dimensionless spin magnitude and $\hat{\mathbf{S}}_\mathbf{i}$ is the unit spin vector. Because the spins of the BHs in a binary system influence the dynamics of the inspiral and merger, a GW detection provides a measurement of the component spins [3].

For an individual GW event, the spin measurements are often poorly constrained [249, 208], but we can combine individual spin posteriors to examine the distribution of dimensionless spin magnitudes across all events. In this Letter, we show that the hierarchical merger scenario yields a unique distribution of BH spin magnitudes a; therefore, by measuring the spins of observed systems, we can constrain this formation process. Our approach is complementary to that of [120], who study the expected distributions of mass, redshift, and binary spin parameter $\chi_{\rm eff}$ for populations of first- and second-generation BHs and show how to use all three measurements to constrain the fraction of second-generation BHs in a detected population. In contrast, we focus solely on GW measurements of spin magnitude a and consider arbitrary generations of previous mergers.

To construct the distribution of BH spin magnitudes resulting from hierarchical mergers, we utilize previous studies of the evolution of BH spins through binary coalescence. Due to advancements in numerical relativity (NR) and post-Newtonian (PN) methods, a number of groups have developed reliable formulae for the final spin following a merger of two spinning BHs [52, 146, 242, 126, 131, 142]. Intuitively, there are two contributions to the spin following a coalescence: the individual spins of the two progenitor BHs and the binary system's orbital angular momentum. As the BBHs inspiral toward each other, they lose energy and orbital angular momentum through the emission of GWs. When the BHs finally merge, as shown by [52], the remaining orbital angular momentum that contributes to the final BH spin can be approximated by the orbital angular momentum of a test particle at the innermost stable circular orbit of the final BH (where the mass of the test particle is taken to be the reduced mass of the BBHs). The contribution from the orbital angular momentum will be most significant for equal mass BBHs and will dominate over the contribution from the spin

angular momentum. For example, as is well understood from NR simulations, a merger of nonspinning BBHs of equal mass will result in a final BH with a dimensionless spin magnitude of 0.6864 [131]. In order for the spins of the BBHs to cancel the orbital angular momentum, resulting in a nonspinning BH, the spins must be sufficiently large and antialigned to the orbital angular momentum, and the mass ratio $q \equiv m_2/m_1 \leq 1$ must be sufficiently small. In fact, using the results of [52], the antialigned contributions to the spins, a_1^-, a_2^-, and the mass ratio, q, must satisfy

$$\frac{1}{q}a_1^- + qa_2^- + 2\sqrt{3} = 0 \tag{2.2}$$

in order to end up with a nonspinning BH. Thus, even for maximally antialigned spins, the mass ratio must satisfy $q < \sqrt{3} - \sqrt{2} \approx 0.32$ in order to overwhelm the orbital angular momentum. As we shall see, this explains why major mergers (in which $q \sim 1$) result in BHs with a relatively high spin distribution, peaked at $a = 0.69$, and with little support below $a \approx 0.5$.

In this work, we consider major mergers ($q \gtrsim 0.7$) as the basis of the hierarchical merger scenario. If the BHs of each generation interact with each other dynamically, they are more likely to form binaries with BHs of similar mass [222, 212] and we would expect mergers of near-equal mass BHs [211, 198]. We would similarly expect near-unity mass ratios for BBHs of primordial origin, as PBH formation scenarios generally allow a narrow mass range for the first generation [159], and we assume that because of dynamical considerations, such BHs only merge with partners of the same generation.

The assumption of major mergers differs from the seminal work of [138], which considered the spin evolution of supermassive BHs as they grow through minor mergers. In contrast to major mergers, minor mergers tend to decrease the spin of the final BH, because the binary's orbital angular momentum is smallest when it augments the total BBH spins (a prograde orbit) and largest when it counteracts it (a retrograde orbit).

We also assume that, in the absence of any aligning mechanism, the spins of each gener-

ation of BHs in the hierarchical merger scenario are isotropically distributed on the sphere. The effects of BBH spins that are preferentially aligned or antialigned with the orbital angular momentum are discussed in Section 2.4. However, it is important to note that spins that are initially partially aligned (antialigned) with the orbital angular momentum can become significantly antialigned (aligned) during the inspiral due to precession [147]. This will not affect an isotropic distribution of spins, as a distribution of spins that is isotropic at large distances will remain isotropic during the inspiral up to the point of plunge [147]. Furthermore, the magnitudes of the BBH spins remain nearly constant during the inspiral (up to 2PN order), which further lends confidence to our calculation of the hierarchical merger spin distribution.

2.3 Methods

2.3.1 Hierarchical Merger Spin Distribution

We apply the formulas of [131] to predict the final BH spin from a merger of two BHs, given the spin vectors and masses of the component BHs. This allows us to build a statistical distribution of spin magnitudes resulting from hierarchical mergers, similar to the distributions found by [242] and [170].

Although we assume major mergers and isotropically distributed spin orientations, we wish to remain general with respect to other aspects of the hierarchical merger scenario. In particular, we do not at the outset specify the spin distribution of the first generation of BHs (before any mergers have occurred) or the exact distribution of mass ratios of merging BHs (although we limit ourselves to $q \geq 0.7$). Furthermore, the desired spin distribution presumably evolves as each generation's BHs merge to form the next generation, but we do not wish to restrict ourselves to a particular generation of the hierarchical merger scenario. Fortunately, as we show below, the resulting spin distribution is relatively insensitive to the

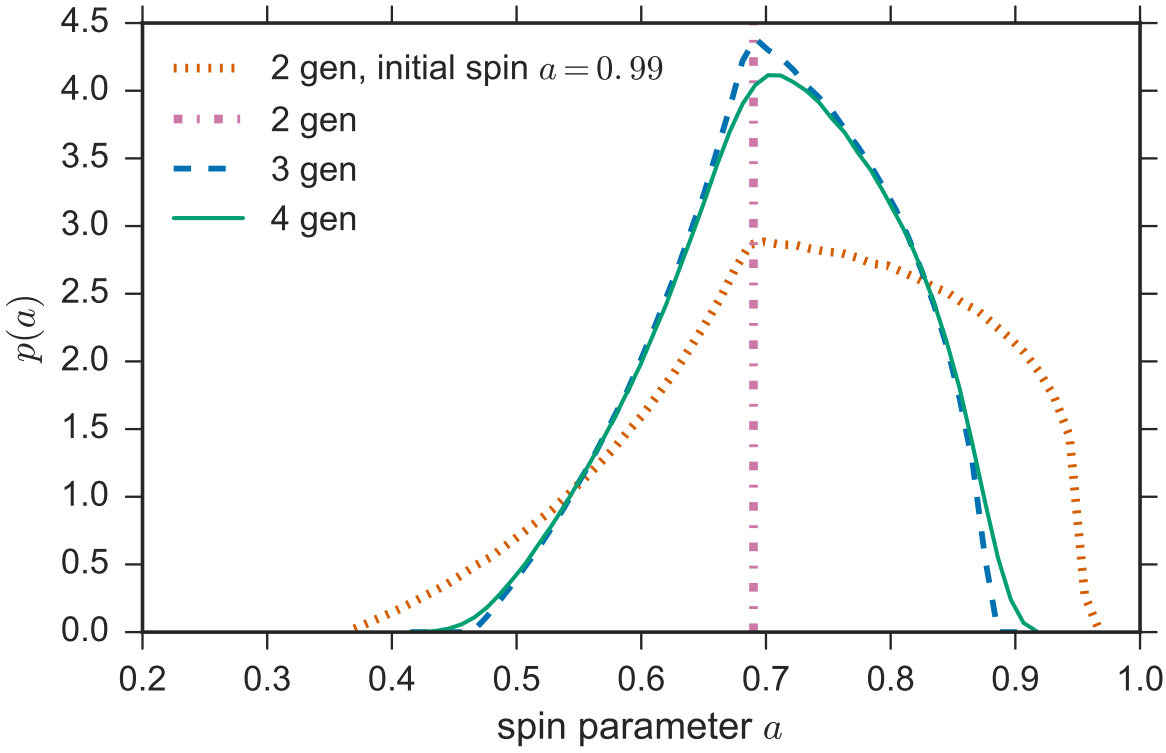

Figure 2.1: **Probability distribution for the dimensionless spin magnitude for each generation of BHs formed through hierarchical mergers.** Unless labeled otherwise, the first generation is nonspinning ($a = 0$) and all mergers take place between equal mass BHs ($q = 1$). For each generation, the spin directions are assumed to be isotropically distributed. Note the rapid convergence to a universal distribution (turquoise solid line). The dotted orange line shows the second-generation distribution for the case where the first generation has near-maximal ($a = 0.99$) spins. The initially nonspinning ($a = 0$) and initially near-maximally spinning ($a = 0.99$) cases are indistinguishable by the fourth generation, converging on the universal distribution.

spin magnitudes of the first generation, the mass ratios (within the range $0.7 \leq q \leq 1$), or which generation we consider (starting with the second generation). We demonstrate this explicitly by computing spin distributions under various choices of these parameters.

We compute probability density functions of dimensionless spin magnitudes as follows: we start by taking a large (6.25×10^6) ensemble of BHs, and then randomly pick pairs of BHs from this first generation and merge each pair, calculating the final spin from the [131] formula. This gives us the distribution of spin magnitudes for the second generation of BHs. In the simplest case we take the first generation of BHs to be all of the same mass and nonspinning, in which case, the second generation's BHs will all be of roughly double the mass and spinning with dimensionless spin magnitude $a = 0.69$. If the initial generation of BHs is equal mass but with isotropic, near-maximal ($a = 0.99$) spins, the second generation of BHs will have a distribution of spin magnitudes that is similarly peaked at $a \sim 0.7$ with slightly wider support (see Fig. 1).

To calculate the spin distribution for the third generation of BHs, we randomly and repeatedly choose pairs of BH spin magnitudes from the second generation and randomly choose their spin directions from a spherically isotropic distribution. This yields the spin magnitudes of the third generation of BHs, and we can iterate this procedure to calculate the distribution of BH spins for the nth generation given the distribution of spins for the $(n-1)$th generation. In practice, we find that the spin distribution changes only slightly between the third and the fourth generation and has fully converged by the fourth generation, regardless of the initial spin distribution (see Fig. 2.2). Different choices of initial spin lead to indistinguishable spin distributions by the third generation.

To explore the effect of different mass ratios, we consider a toy model in which all mergers occur with a mass ratio of $q = 0.7$, instead of $q = 1$ as assumed above. We find that the resulting spin distributions are very similar (see Fig. 2.3.1), suggesting that any distribution of mass ratios in the range $0.7 \leq q \leq 1$ (as expected for dynamically forming binaries)

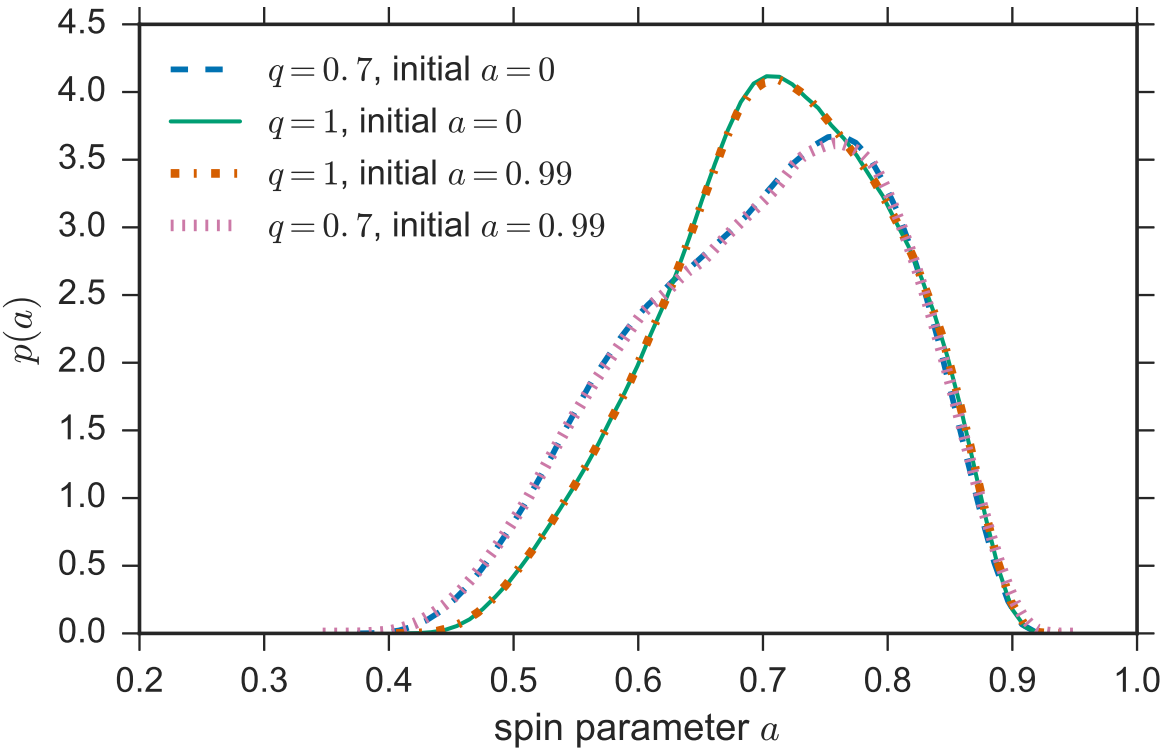

Figure 2.2: **Converged spin distributions (fourth generation) of hierarchically formed BHs**, where in one scenario the BHs always merge with equal mass ($q = 1$) and in the other scenario the BHs always merge with mass ratio $q = 0.7$. Changing the spin of the first generation leads to indistinguishable distributions.

would not significantly affect the distribution of final spins. Thus, we find that regardless of mass ratio and initial spin, the hierarchical merger scenario gives rise to a single, standard distribution of BH dimensionless spin magnitudes sharply peaked at $a \sim 0.7$ with nonzero support over $0.4 \lesssim a \lesssim 0.9$. In what follows when we refer to the hierarchical merger spin distribution, we mean the $q = 1$ distribution shown in Fig. 2.3.1, calculated as the fourth generation of equal mass mergers. An alternate choice would not significantly affect our results.

Our findings are consistent with [47], who found that following a single merger the final spin magnitude is $a \approx 0.7$ regardless of the initial spin magnitude, assuming we average over an isotropic distribution of spin directions (see their Fig. 2). Our results are also consistent with [242] and [170], both of whom found that the distribution of spin magnitudes converges after four generations of repeated mergers. However, note that by limiting ourselves to the major mergers relevant for stellar mass BHs, our hierarchical merger spin distribution is different from the distribution presented in Fig. 19 of [170], as they considered a wide distribution of mass ratios appropriate for supermassive BHs. In particular, our distribution has little support below $a \sim 0.5$. Our hierarchical merger spin distribution is most similar to the $q = 1$ distribution in Fig. 1 of [242] and Fig. 20 of [170]; however, we argue that this distribution is insensitive to the initial BH spins and is an adequate description of the second and third generation of BHs even before it fully converges in the fourth generation.

2.3.2 Mixture Model Analysis

We apply a hierarchical Bayesian framework [132, 180] to analyze a collection of BH spin measurements, where the spin measurement from the ith GW detection takes the form of a two-dimensional posterior for the BBH spin magnitudes $p(\boldsymbol{\alpha}_i|d_i)$, where $\boldsymbol{\alpha}_i = (a_{1,i}, a_{2,i})$ and d_i is the data. To understand the true population of BH spins from the observed BBHs, we assume that the true spin distribution is parameterized in terms of some parameters, $\mathbf{A}$,

which we seek to infer. The true spin distribution, therefore, can be written as $p(\boldsymbol{\alpha}|\mathbf{A})$, and we want to know $p(\mathbf{A}|\mathbf{d})$, where $\mathbf{d} = \{d_i\}$ is the data across all GW detections. We assume that each GW detection is independent, so that

$$p(\mathbf{A}|\mathbf{d}) = \prod_i p(\mathbf{A}|d_i). \tag{2.3}$$

Furthermore, we have

$$p(\mathbf{A}|d_i) = \int d\boldsymbol{\alpha}_i p(\mathbf{A}, \boldsymbol{\alpha}_i|d_i), \tag{2.4}$$

and applying Bayes's rule gives

$$p(\mathbf{A}, \boldsymbol{\alpha}_i|d_i) \propto p(d_i|\boldsymbol{\alpha}_i)p(\boldsymbol{\alpha}_i|\mathbf{A})p(\mathbf{A}), \tag{2.5}$$

where $p(d_i|\boldsymbol{\alpha}_i)$ is the two-dimensional likelihood for the BBH spins, and $p(\mathbf{A})$ is the prior probability for the population parameters $\mathbf{A}$. Before we have learned anything about the population distribution of spin magnitudes $p(\boldsymbol{\alpha}_i|\mathbf{A})$, in the analysis of individual events, we assume a two-dimensional flat prior on $\boldsymbol{\alpha}_i$, so that the likelihood $p(d_i|\boldsymbol{\alpha}_i)$ is proportional to the posterior $p(\boldsymbol{\alpha}_i|d_i)$. Putting together equations 2.3–2.5, we have

$$p(\mathbf{A}|\mathbf{d}) \propto \prod_i \left[\int d\boldsymbol{\alpha}_i p(d_i|\boldsymbol{\alpha}_i)p(\boldsymbol{\alpha}_i|\mathbf{A}) \right] p(\mathbf{A}) \tag{2.6}$$

where, because $p(d_i|\boldsymbol{\alpha}_i) \propto p(\boldsymbol{\alpha}_i|d_i)$ for a single event, we can evaluate the above integral over $\boldsymbol{\alpha}_i$ of $p(\boldsymbol{\alpha}_i|\mathbf{A})$ weighed by the likelihood $p(d_i|\boldsymbol{\alpha}_i)$ as an average over N_i posterior samples $\boldsymbol{\alpha}_i^k$:

$$\begin{aligned} \int d\boldsymbol{\alpha}_i p(d_i|\boldsymbol{\alpha}_i)p(\boldsymbol{\alpha}_i|\mathbf{A}) &= \langle p(\boldsymbol{\alpha}_i|\mathbf{A}) \rangle_{\boldsymbol{\alpha}_i} \\ &\approx \frac{1}{N_i} \sum_{k=1}^{N_i} p(\boldsymbol{\alpha}_i^k|\mathbf{A}). \end{aligned} \tag{2.7}$$

In our case, in order to investigate whether the detected BBHs favor the hierarchical

merger scenario, we write the true spin population as a mixture model. Lacking a strong astrophysical prior on the distribution of BH spins [191], we take some fraction f_u of the BHs to be uniformly spinning over the allowed range $[0, 1]$, and the remaining $(1 - f_u)$ of the BHs to come from the hierarchical merger population. It is straightforward to consider alternate spin magnitude distributions, and the same analysis would apply if we included an additional component in the mixture model or replaced the uniformly distributed component with a different spin distribution. For the mixture model with parameter f_u, we have

$$p(a|f_u) = \begin{cases} f_u + (1 - f_u)p_{\rm hm}(a) & 0 \leq a \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.8}$$

where $p_{\rm hm}(a)$ is the hierarchical merger spin distribution. We assume the spins of the BBHs in a single system are independent of one another, so

$$p(\boldsymbol{\alpha}_i|f_u) = p(a_{1,i}|f_u)p(a_{2,i}|f_u). \tag{2.9}$$

We also use a flat prior for the mixture parameter

$$p(f_u) = \begin{cases} 1 & 0 \leq f_u \leq 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.10}$$

Then for the mixture model, equation 2.6 becomes

$$\begin{aligned} p(f_u|\mathbf{d}) \propto p(f_u) \prod_i \int d\boldsymbol{\alpha}_i p(d_i|\boldsymbol{\alpha}_i) \left[f_u + (1 - f_u)p_{\rm hm}(a_{1,i})\right] \\ \cdot \left[f_u + (1 - f_u)p_{\rm hm}(a_{2,i})\right] \end{aligned} \tag{2.11}$$

Using equation 2.7, given N_i posterior samples for each BBH, we can approximate equation

2.11 as

$$p(f_u|\mathbf{d}) \propto p(f_u)\prod_i f_u^2 + f_u(1-f_u)\frac{1}{N_i}\sum_{k=1}^{N_i}\left[p_{\rm hm}(a^k_{1,i}) + p_{\rm hm}(a^k_{2,i})\right] + (1-f_u)^2\frac{1}{N_i}\sum_{k=1}^{N_i} p_{\rm hm}(a^k_{1,i})p_{\rm hm}(a^k_{2,i}). \tag{2.12}$$

The mixture model parameterization provides a convenient way to compare the hierarchical merger model to any other model (in our case, a model that yields a flat distribution of spins). As [252] discuss, for a mixture of two (or more) models, we can write the posterior of the mixture parameter in terms of the Bayesian evidence for the models under consideration. In our case, we can relate the posterior $p(f_u|\mathbf{d})$ to the evidence ratio, or Bayes factor, between the hierarchical merger model $\mathcal{H}_{\rm hm}$ and the uniform spin model $\mathcal{H}_u$. The evidence for each model given data d_i is defined as $Z^i_{\rm hm} = p(d_i|\mathcal{H}_{\rm hm})$ and $Z^i_u = p(d_i|\mathcal{H}_u)$, and, assuming GW detections are independent, we can write

$$\begin{aligned}
p(f_u|\mathbf{d}) &\propto p(\mathbf{d}|f_u)p(f_u) && (2.13)\\
&= p(f_u)\prod_i p(d_i|f_u) && (2.14)\\
&= p(f_u)\prod_i \left[p(d_i,\mathcal{H}_u|f_u) + p(d_i,\mathcal{H}_{\rm hm}|f_u)\right] && (2.15)\\
&= p(f_u)\prod_i \left[p(d_i|\mathcal{H}_u)p(\mathcal{H}_u|f_u) + p(d_i|\mathcal{H}_{\rm hm})p(\mathcal{H}_{\rm hm}|f_u)\right] && (2.16)\\
&= p(f_u)\prod_i \left[Z^i_u f_u + Z^i_{\rm hm}(1-f_u)\right]. && (2.17)
\end{aligned}$$

We therefore have that the Bayes's factor is

$$\prod_i \frac{Z^i_u}{Z^i_{\rm hm}} = \frac{p(f_u=0)}{p(f_u=1)}\frac{p(f_u=1|\mathbf{d})}{p(f_u=0|\mathbf{d})}. \tag{2.18}$$

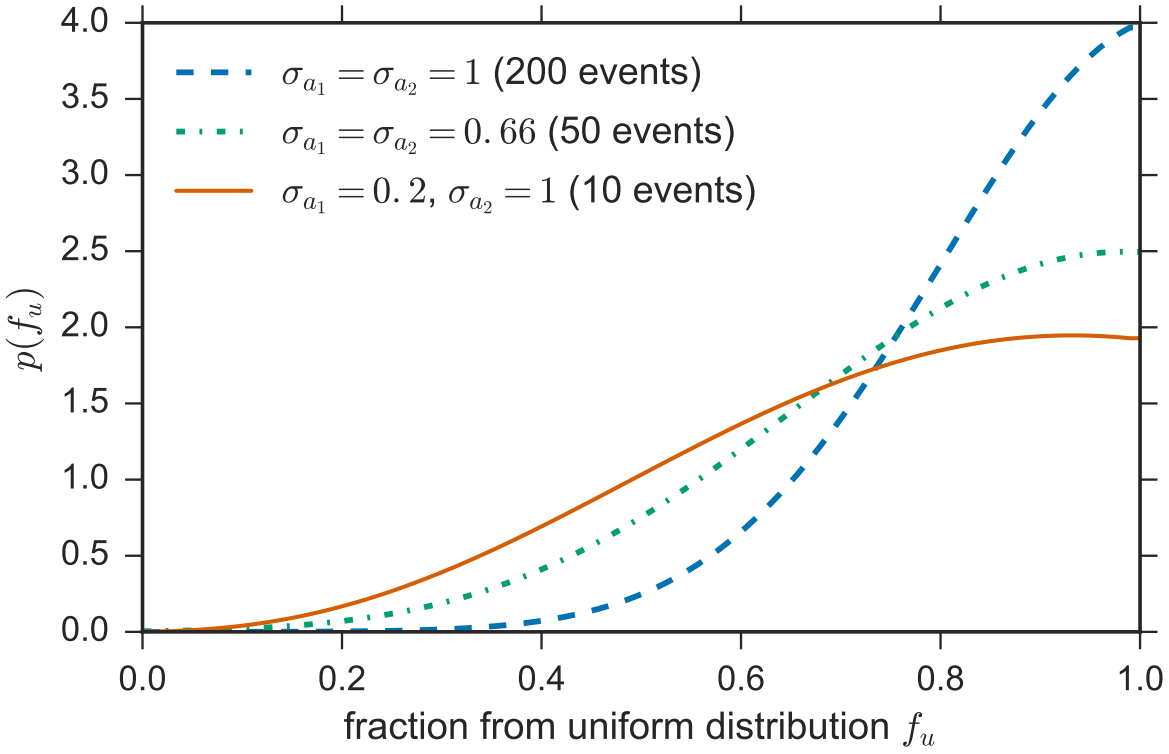

Figure 2.3: **Posterior probability density functions on the parameter f_u for 3 simulated populations of BHs that have not formed hierarchically** (the true $f_u = 1$). The populations of BHs all have spin magnitudes drawn from a uniform $[0, 1]$ distribution, but differ in the uncertainty on their measured spins magnitudes. We approximate spin measurements as truncated Gaussians and vary σ_{a_1}, σ_{a_2} between populations. We note that 10 detections may be sufficient to rule out a pure hierarchical merger model.

Thus, computing $p(f_u|\mathbf{d})$ allows us to directly find the Bayes's factor, which allows us to argue (or refute) that a population of observed BHs came from the hierarchical merger formation channel. The mixture model is also useful to constrain the fraction of the observations that are consistent with having formed through hierarchical mergers. In the next section, we demonstrate that this analysis will yield meaningful constraints within just a few years of advanced LIGO operation.

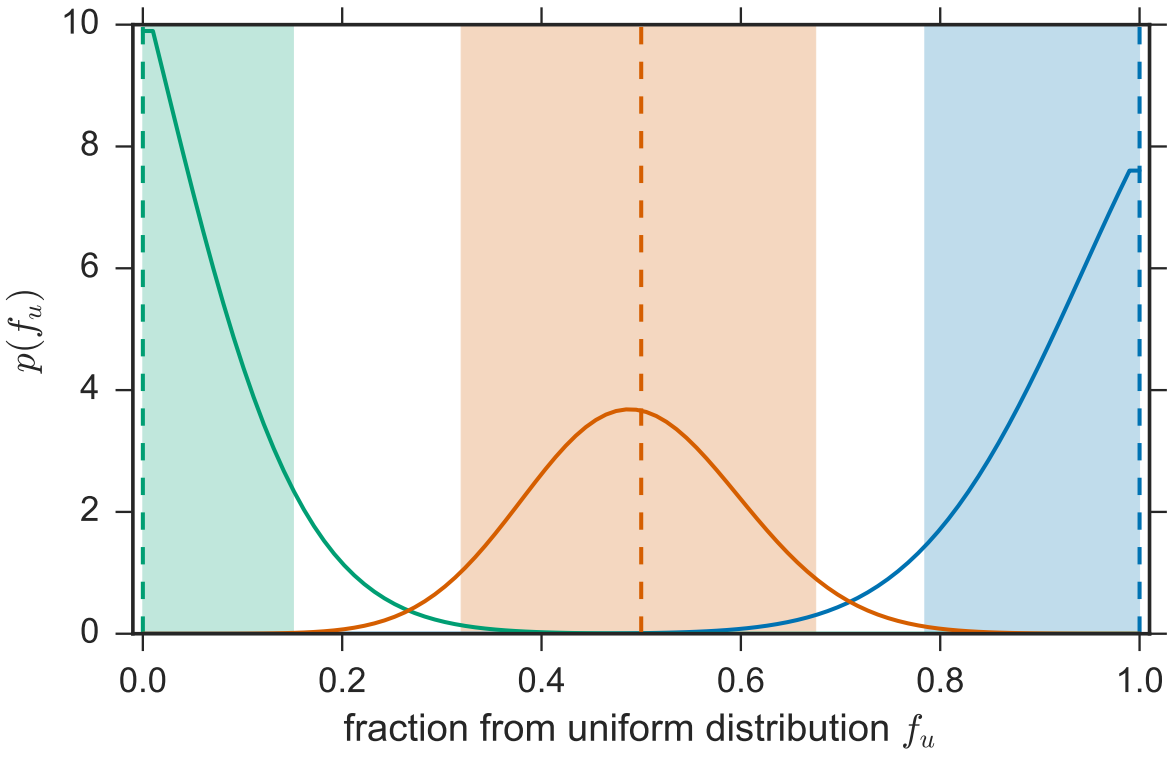

Figure 2.4: **Posterior probability density functions on the parameter f_u for 3 simulated populations with different values of f_u** (given by the dashed lines). The simulated BH populations with $f_u = 0$ (turquoise) and $f_u = 1$ (sky blue) each consist of 200 BBH events, while the simulated BH population with $f_u = 0.5$ (orange) consists of 400 BBH events. We assume that for all events the spin magnitude measurements $p(a_1|d_i)$, $p(a_2|d_i)$ are truncated Gaussians with standard deviations $\sigma = 0.66$. The shaded regions denote 90% credible intervals.

2.4 Results

For the purposes of this work, we illustrate our method on very simplified spin posterior distributions, leaving the analysis of real data to the LIGO collaboration. We assume that each BBH detection provides a measurement of the two component BH spin magnitudes with some uncertainty [see, for example, Fig. 5 of 3]. We neglect correlations between the two spin measurements, which is equivalent to setting

$$p(d_i|\boldsymbol{\alpha}_i) = p(d_i|a_{1,i})p(d_i|a_{2,i}) \tag{2.19}$$

in equation 2.11. Following [236], we approximate each spin magnitude posterior $p(a_{j,i}|d_i)$ ($j = \{1, 2\}$) as a Gaussian, restricted to the range [0,1], with a standard deviation $\sigma_{j,i}$ corresponding to the measurement uncertainty. In other words, for a BH with dimensionless spin magnitude $a_{j,i}^{\rm true}$, we generate a posterior centered on $a_{j,i}^{\rm data} = a_{j,i}^{\rm true} + e_{j,i}$ where $e_{j,i}$ is a random measurement error chosen from $N(0, \sigma_{j,i})$. The spin magnitude posterior for a single BH is then given by

$$p(a_{j,i}|d_i) = N(a_{j,i}^{\rm data}, \sigma_{j,i}) \tag{2.20}$$

truncated and normalized to our prior range $[0, 1]$. With these assumptions, we compute equation 2.11 by drawing 1000 samples from each spin magnitude posterior given by equation 2.20 for a simulated population of $a_{j,i}^{\rm true}$. In other words, we solve

$$p(f_u|\mathbf{d}) \propto p(f_u) \prod_i \prod_{j=1,2} f_u + (1 - f_u)\langle p_{\rm hm}(a_{j,i})\rangle_{a_{j,i}} \tag{2.21}$$

where

$$\langle p_{\rm hm}(a_{j,i})\rangle_{a_{j,i}} \approx \frac{1}{1000} \sum_{k=1}^{1000} p_{\rm hm}(a_{j,i}^k) \tag{2.22}$$

for $a_{j,i}^k \sim p(a_{j,i}|d_i)$.

The uncertainty σ on spin magnitude depends on various factors, including the true spin magnitudes, the signal-to-noise ratio (SNR) of the inspiral and ringdown, the mass ratio of the binary, and the orientation of the spin vectors. As demonstrated by [208], we expect this uncertainty to be rather large and not particularly dependent on the SNR, especially for events where there is little power in the ringdown, so we carry out our analysis with the conservative choice of $\sigma = 1$. We then repeat the analysis under the assumption that all events are like GW150914 in terms of spin magnitude uncertainty: consisting of one relatively well measured BH spin magnitude with $\sigma = 0.2$, and one poorly measured BH spin magnitude with $\sigma = 1$. This can be expected for events with moderately high SNR in both the inspiral (which constrains the weighted aligned spin combination $\chi_{\rm eff}$) and the ringdown (which constrains the spin of the final BH a_f). Motivated by the choice of posterior uncertainties in [236] and an examination of mass ratio uncertainties and covariances for published LIGO events [see Table 1 and Fig. 4 of 3], we repeat our analysis for spin posteriors with $\sigma = 0.66$.

Our results are similar for all choices of spin magnitude uncertainties, suggesting that LIGO will be able to clearly distinguish between a population of hierarchically formed BHs and a population of uniformly spinning BHs with $\mathcal{O}(100)$ detections (see Fig. 2.4), although this will be possible with as few as 10 detections if at least one spin component is relatively well-measured ($\sigma \approx 0.2$) as in the case of GW150914 (see Fig. 2.3.2). If the true population of detected BHs is mixed, it requires more detections to precisely measure the fraction $1 - f_u$ that have spin magnitudes consistent with formation through hierarchical mergers. We see in Fig. 2.4 that the 90% confidence interval for f_u is relatively wide for a mixed population with 400 events, although if we are simply interested in ruling out $f_u = 0$ or $f_u = 1$, $\mathcal{O}(100)$ detections is sufficient even in the most pessimistic case considered. It is straightforward to extrapolate these results to a greater number of detections: as expected, the width of the posterior $p(f_u|\mathbf{d})$ decreases with the number of detections, N, as $1/\sqrt{N}$.

We have thus far analyzed the characteristic spin distribution resulting from major

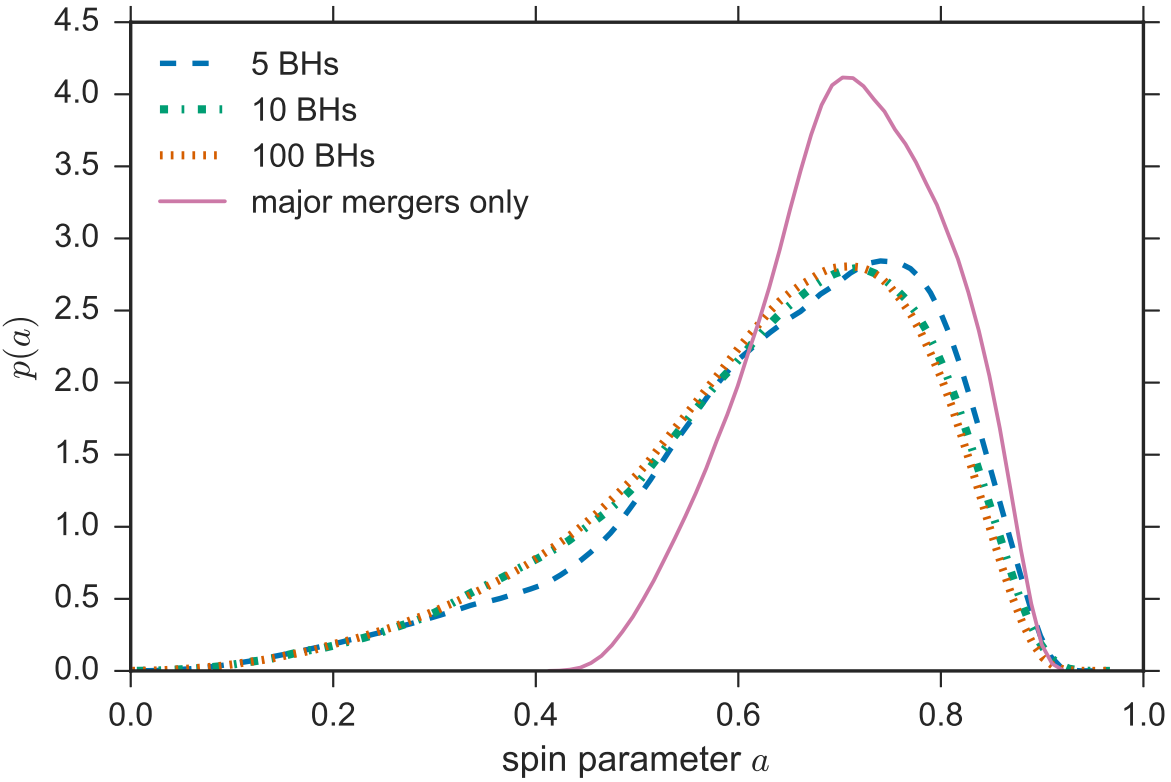

Figure 2.5: **Probability density function of the spin magnitude of a final BH formed through a "cluster catastrophe"** of a fixed number of BHs, compared to the universal hierarchical merger distribution. Initially the N BHs are of equal mass and spin magnitudes drawn uniformly on $[0, 1]$, and they merge with each other in randomly selected pairs (with isotropic spin directions) until they have all merged into a single final BH.

($q \geq 0.7$) mergers of isotropically spinning BHs. Below, we discuss the implications of relaxing these assumptions. We might expect deviations from isotropic spins in certain astrophysical situations such as BBH formation in a gas-rich AGN disk, where the spins of the component BHs can be preferentially aligned or antialigned with the orbital angular momentum [187]. If the spins of BBHs are always aligned with the orbital angular momentum, it is straightforward to see that this will result in an even narrower spin distribution, strongly peaked at very high spin magnitudes, converging to $a \sim 0.9$ by the third generation. This situation will thus be easier to constrain with LIGO data. If both component spins are always antialigned with the orbital angular momentum and the first generation has moderately high spins $a \sim 0.7$, the future generations will have spins $0.43 < a < 0.54$. Even assuming maximal antialigned initial spins, while the second-generation products will be spinning at $0.32 < a < 0.36$ for mass ratios $q \geq 0.7$, the spin distribution will converge to $a \sim 0.5$ starting with the third generation. If the spins are equal in magnitude but one is aligned and the other antialigned with the orbital angular momentum, the situation is identical to a merger of nonspinning BHs and yields a spin magnitude $a \sim 0.7$ (although note that the best-constrained spin parameter $\chi_{\rm eff}$ for the binary will be $\chi_{\rm eff} = 0$). We conclude that hierarchical major mergers of BBHs cannot produce low spin magnitudes ($a \lesssim 0.4$).

We can relax the assumption of major mergers by considering an alternative "cluster catastrophe" formation scenario in which a fixed number, N, of equal mass BHs repeatedly merge in randomly chosen pairs, irrespective of the mass ratio, until there is a single remaining BH. We take the initial distribution of spin magnitudes to be uniform in $[0, 1]$ and spin directions to always be isotropic. We find that the spin magnitude of the single remaining BH is insensitive to N or the initial spin magnitudes and is distributed according to the probability distribution in Fig. 2.4. While such a scenario can lead to low mass-ratio mergers in which the primary spin may cancel the orbital angular momentum and produce spin magnitudes $a < 0.4$, these low spin magnitudes remain unlikely: spin magnitudes $a < 0.4$

are produced less than 10% of the time.

2.5 Conclusion

We have shown that if BHs build up through hierarchical major mergers of smaller BHs, the spin magnitudes of the resulting BHs follow a universal distribution (see Fig. 2.3.1). Because this distribution is relatively independent of the details of the hierarchical merger scenario, we can use it to test whether an observed population of BHs was formed through hierarchical mergers. Although a GW observation of a single coalescing BBH does not strongly constrain individual BH spins [3], we estimate that the hierarchical formation channel will be strongly constrained with $\mathcal{O}(100)$ detections. Furthermore, we have shown that hierarchical mergers rarely produce BHs with spins below $a \lesssim 0.5$, and even in extreme scenarios that favor antialigned spins or result in a "cluster catastrophe," one rarely finds BH spins below $a \lesssim 0.4$. If BHs do not form through hierarchical mergers and the spin distribution is uniform on [0,1], we have shown that it will be possible to rule out the hierarchical formation channel with a sample of $\mathcal{O}(10)$ detections (see Fig. 2.3.2). If instead we select between a hierarchical model and one that favors low spins (instead of uniform as done above), even fewer detections would be sufficient to falsify either model. We note that the spin constraints of the primary component BHs of GW150914 and LVT151012 appear to favor spins $a \lesssim 0.5$ over spins $a \sim 0.7$ (see Figs. 5 of 6, 3), suggesting that they are unlikely to have formed through hierarchical major mergers. We leave a quantitative analysis of these events for future work.

CHAPTER 3

WHERE ARE LIGO'S BIG BLACK HOLES?

The first few binary mergers detected by LIGO and Virgo had component black holes in the mass range $\sim$ 10–30 $M_\odot$. While these component black holes are massive compared to the electromagnetically-detected stellar mass black holes (with masses in the range 5–20 $M_\odot$), they are not so massive when we consider that LIGO is sensitive to mergers involving black holes at least as massive as 100 $M_\odot$. In fact, at a fixed luminosity distance, the amplitude of a gravitational-wave signal increases as its chirp mass $\mathcal{M}_c^{5/6}$, meaning that the more massive the binary, the farther out LIGO can detect it. LIGO's sensitivity grows as a function of binary mass up to total masses of roughly 200 $M_\odot$; after this, the merger takes place at frequencies too low for LIGO's sensitive band (between $\sim$ 10–1000 Hz).

In this chapter, we find that the sensitive spacetime volume probed by LIGO grows with the total mass M of the binary roughly as $M^{2.2}$. Therefore, if the black holes follow a similar power-law distribution to the Salpeter initial mass function for stars, $M^{-2.3}$, we expect that the *detected* distribution of black holes will be nearly flat across the entire mass range out to component masses of $\sim$ 100 $M_\odot$. However, there have been no detections involving black holes in the mass range $\sim$ 40–100 $M_\odot$, suggesting a physical dearth of black holes in this mass range. We fit a power-law model to the black hole mass distribution, allowing both the slope and the maximum mass of the power-law to vary, and infer a maximum mass at $\sim$ 40 $M_\odot$ (with $m_{\rm max} < 77$ $M_\odot$ at 95% credibility). The evidence for a maximum mass at $\sim$ 40 $M_\odot$ is remarkably coincident with predictions from supernova theory, which predict that the onset of pair-production in stellar cores between $\sim$ 40 and 120 solar masses lead to unstable supernova explosions which prevent the formation of black holes with masses in this range. Directly measuring the mass spectrum of black holes with gravitational wave detections allows us to test predictions of massive star evolution and supernova explosion physics. Additionally, by testing whether binary black holes respect the "pair-instability

gap" starting at $\sim 40\ M_\odot$, we are probing whether these systems are formed from stars, as opposed to primordial black holes or perhaps from mergers of smaller black holes (see Chapter 1).

This chapter was written in collaboration with Daniel Holz, and published in the Astrophysical Journal Letters [105]. The original analysis was based on only the first four LIGO detections. However, I later performed this analysis on the ten binary black hole detections from LIGO-Virgo's first and second observing runs, as part of the LVC paper [19]. With the full dataset, we inferred even stronger constraints on the maximum mass, measuring $m_{\rm max} = 42^{+15}_{-6}\ M_\odot$. We later realized that this feature in the mass distribution could also be used to measure the expansion history of the universe; see [98].

MF was supported by the NSF Graduate Research Fellowship Program under Grant No. 1746045. MF and DEH were partially supported by NSF CAREER grant PHY-1151836 and NSF grant PHY-1708081. They were also supported by the Kavli Institute for Cosmological Physics at the University of Chicago through NSF grant PHY-1125897 and an endowment from the Kavli Foundation. DEH thanks the Niels Bohr Institute for its hospitality while part of this work was completed, and acknowledges the Kavli Foundation and the DNRF for supporting the 2017 Kavli Summer Program.

3.1 Abstract

In LIGO's O1 and O2 observational runs, the detectors were sensitive to stellar mass binary black hole coalescences with component masses up to $100\,M_\odot$, with binaries with primary masses above $40\,M_\odot$ representing $\gtrsim 90\%$ of the total accessible sensitive volume. Nonetheless, of the 5.9 detections (GW150914, LVT151012, GW151226, GW170104, GW170608, GW170814) reported by LIGO-Virgo, the most massive binary detected was GW150914 with a primary component mass of $\sim 36\,M_\odot$, far below the detection mass limit. Furthermore, there are theoretical arguments in favor of an upper mass gap, predicting an absence

of black holes in the mass range $50 \lesssim M \lesssim 135\,M_\odot$. We argue that the absence of detected binary systems with component masses heavier than $\sim 40\,M_\odot$ may be preliminary evidence for this upper mass gap. By allowing for the presence of a mass gap, we find weaker constraints on the shape of the underlying mass distribution of binary black holes. We fit a power-law distribution with an upper mass cutoff to real and simulated BBH mass measurements, finding that the first 3.9 BBHs favor shallow power law slopes $\alpha \lesssim 3$ and an upper mass cutoff $M_{\rm max} \sim 40\,M_\odot$. This inferred distribution is entirely consistent with the two recently reported detections, GW170608 and GW170814. We show that with ~ 10 additional LIGO-Virgo BBH detections, fitting the BH mass distribution will provide strong evidence for an upper mass gap if one exists.

3.2 Introduction

One of the most fundamental quantities in gravitational-wave astrophysics is the mass distribution of stellar-mass black holes (BHs). Characterizing this distribution in merging binary systems is crucial to understanding stellar evolution, supernova physics, and the formation of compact binary systems. Prior to the first gravitational-wave (GW) detections of binary black holes (BBHs), the sample of ~ 20 BHs in X-ray binary systems was used to infer the BH mass distribution [200, 97], providing strong evidence for the existence of a mass gap between the heaviest neutron star (NS) (~ 2–$3\,M_\odot$) and the lightest BH (~ 4–$5\,M_\odot$) [but see also 160]. The presence of a mass gap between NSs and BHs has critical implications for supernova explosion theory [42], and there are several proposed methods to probe this mass gap with gravitational-wave observations of compact binaries [168, 175, 159]. In addition to the low-mass gap, supernova theory suggests that pulsational pair-instability supernovae (PPISN) [127] and pair instability supernovae (PISN) [115, 209, 49] lead to a second mass gap between ~ 50 and $135\,M_\odot$ for BHs formed from stellar core collapse [37, 185, 38, 254, 231]. Although studies of the lower mass gap have to wait for many more binary detections, be-

cause LIGO's sensitivity is almost 500 times greater for 50–50 $M_\odot$ mergers than 3–3 $M_\odot$ mergers, the existing data already begins to probe the upper mass gap.

For the first four BBH detections (GW150914, LVT151012, GW151226, and GW170104), the LIGO-Virgo collaboration fit the BBH mass distribution with a power law parametrization on the primary BBH mass, m_1, [3, 11] inspired by the stellar initial mass function (IMF) [216, 162]. Specifically, [3] use the following one-parameter power law to model the distribution of primary component BH masses:

$$p\left(m_1 \mid \alpha\right) \propto m_1^{-\alpha}, \tag{3.1}$$

where $M_{\rm min} < m_1 < M_{\rm max}$. The mass ratio between component BHs, $q \equiv m_2/m_1 \leq 1$, is assumed to be uniformly distributed in the allowed range $M_{\rm min}/m_1 \leq q \leq \min(M_{\rm tot,max}/m_1 - 1, 1)$. Thus, the marginal distribution of the secondary component mass, m_2, is given by:

$$p\left(m_2 \mid m_1\right) \propto \frac{1}{\min(m_1, M_{\rm tot,max} - m_1) - M_{\rm min}}, \tag{3.2}$$

and therefore the joint mass distribution is:

$$p\left(m_1, m_2 \mid \alpha\right) \propto \frac{m_1^{-\alpha}}{\min(m_1, M_{\rm tot,max} - m_1) - M_{\rm min}}. \tag{3.3}$$

The only free parameter in this assumed mass distribution is the power law slope, α. The minimum BH mass, $M_{\rm min}$, is fixed at $M_{\rm min} = 5\,M_\odot$ and the maximum mass, $M_{\rm max}$, is fixed at $M_{\rm max} = 100\,M_\odot - M_{\rm min}$. Meanwhile, the total BBH mass, $M_{\rm tot} = m_1 + m_2$, is also restricted: $M_{\rm tot} \leq M_{\rm tot,max} = 100\,M_\odot$. (Enforcing $M_{\rm tot,max} = 100\,M_\odot$ causes a break in the power law at $50\,M_\odot$.) The choice for $M_{\rm min}$ is motivated by the empirical lower mass gap, while $M_{\rm max}$ and $M_{\rm tot,max}$ are set by the stellar binary matched-filter search, which defines stellar-mass BBHs as those with source-frame total masses $m_1 + m_2 \leq 100\,M_\odot$ [57, 243, 4].

However, LIGO is in principle sensitive to heavier BBHs [16]. BBHs with detector-frame total masses up to $600\,M_\odot$ can be detected via matched-filtering by the intermediate mass black hole (IMBH) modeled search [195, 189, 77], and IMBHs of even higher mass can be detected by the unmodeled transient search [153, 10].

We observe that a key assumption of the distribution in Eq. 3.3 is that BHs in merging binaries follow the same mass distribution from $5\,M_\odot$ to at least $50\,M_\odot$, and that there exist BHs as heavy as $95\,M_\odot$. Meanwhile, LIGO is extremely sensitive to heavy BBHs. The first-order post-Newtonian approximation predicts that for low mass BBHs and a Euclidean universe, the spacetime volume, VT, to which LIGO can detect a BBH merger of a fixed mass ratio increases with its primary component mass, m_1, roughly as $VT \propto m_1^{5/2}$. In the following section, we find that when accounting for cosmology and taking BBHs over the entire range $10\,M_\odot < M_{\rm tot} < 100\,M_\odot$, it is still a good approximation to take $VT \propto m_1^k$, with $k \sim 2.2$. This means that if the BBH mass distribution follows a power law with slope α as in Eq. 3.3, we expect the mass distribution among detected BHs to follow $m_1^{-\alpha+2.2}$. For a Salpeter IMF ($\alpha = 2.35$), this implies an almost flat *detected* distribution of binary black hole masses. Thus, the absence of heavy BBHs in the data quickly indicates either that the mass distribution declines steeply towards high masses ($\alpha \gg 2.2$), or that an upper mass gap sharply cuts off the mass distribution.

In this Letter we show that we can start to distinguish between these two scenarios with the first four LIGO BBH detections (including LVT151012, which has an 87% probability of being astrophysical) [3, 11]. Using simulated BBH detections, we demonstrate that if an existing mass gap is not accounted for, the non-detection of heavy stellar mass BHs will quickly bias the power law fit to distributions which are erroneously too steep. However, by including a maximum BH mass as a free parameter in the analysis, we can simultaneously infer the shape of the mass distribution and the location of a mass gap, if present. We carry out this analysis for the first four BBHs as well as for simulated BBH detections.

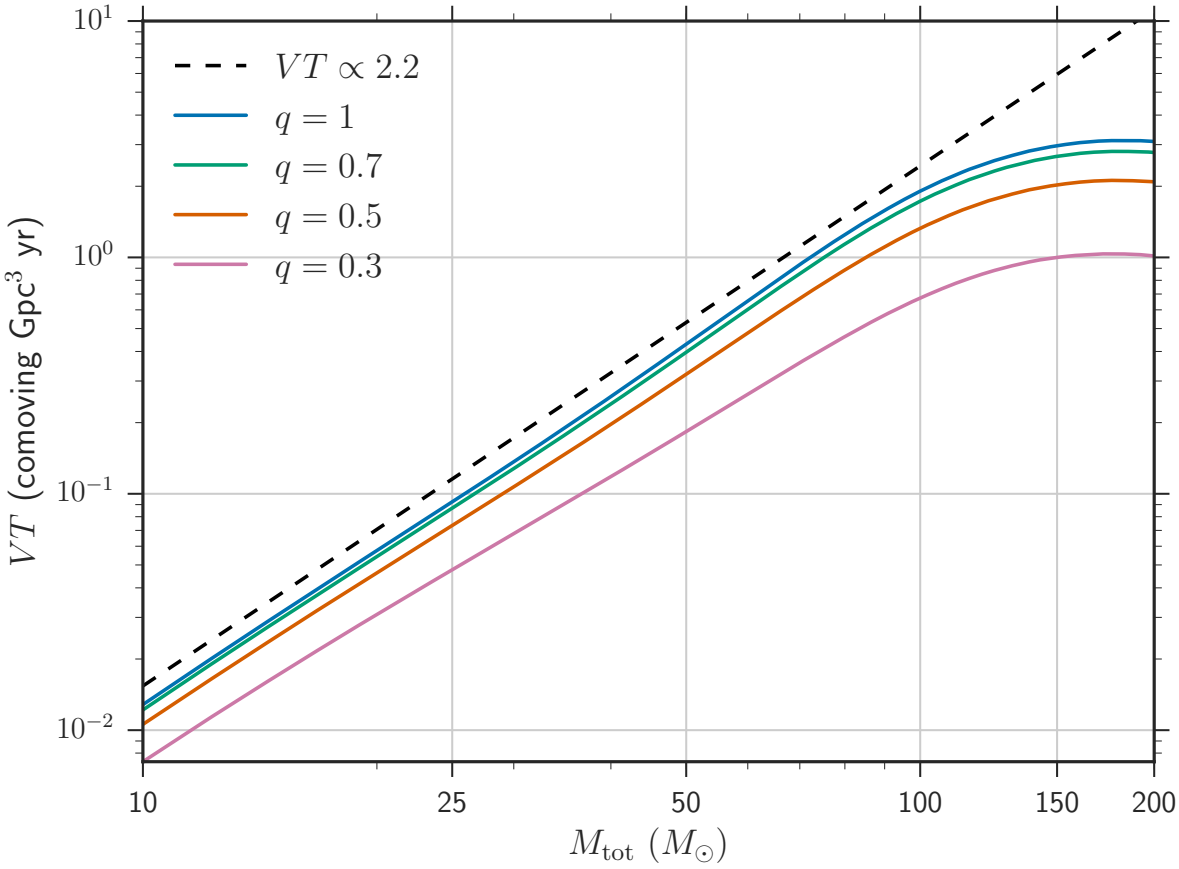

Figure 3.1: **Sensitive redshifted spacetime volume**, VT, of the LIGO detectors in O1 and O2 as a function of BBH total mass, $M_{\rm tot}$, and mass ratio, q, calculated under the semi-analytic approximation described in the text for one year of observation. We find that $VT \propto m_1^{2.2}$ over the range $10\,M_\odot \leq M_{\rm tot} \leq 100\,M_\odot$.

We find that there is already evidence for an upper mass cutoff at $\sim 40\,M_\odot$ from the first four detections, a conclusion that is further supported by the two recently reported BBH detections (GW170608 and GW170814) [13, 12]. We show that with $\mathcal{O}(10)$ additional detections, the presence and location of the bottom edge of the mass gap will be highly constrained.

3.3 Sensitive volume

As we noted in the previous section, for a given mass ratio, the sensitivity of the LIGO-Virgo search scales with primary component mass roughly as ${m_1}^{2.2}$. We characterize the sensitivity by the redshifted spacetime volume, VT, for which a given search is sensitive to a BBH system of given masses. If we assume that the rate of BBH coalescences is uniform in comoving volume and source-frame time and neglect the effects of BH spin on the detectability of a source, VT depends only on the power spectral density (PSD) curve characterizing the detectors and the BBH component masses [8]. Under these assumptions, VT is given by:

$$VT(m_1, m_2) = T \int dz \frac{dV_c}{dz} \frac{1}{1+z} f(z, m_1, m_2), \tag{3.4}$$

where T is the search time, V_c is the comoving volume, and $0 < f(z, m_1, m_2) < 1$ is the probability that a BBH system of masses m_1, m_2 at redshift z will be detected. We adopt the cosmological parameters from [24] throughout the calculation. To calculate the detection probability, $f(z, m_1, m_2)$, we use the semi-analytic approximation from [8]. Taking the PSD function corresponding to the early aLIGO high-sensitivity scenario in [7] (a good approximation to the PSD during the first and second aLIGO observing runs), we calculate the optimal matched-filter signal-to-noise ratio (SNR) of a BBH with component masses m_1 and m_2 and zero spins located at redshift z. The optimal SNR, $\rho_{\rm opt}$, corresponds to a face-on source that is directly overhead to a single detector. We then generate random angular factors, $0 < w < 1$, from a single-detector antenna power pattern, assuming that binaries are distributed uniformly on the sky with isotropic inclination vectors [101, 87, 41]. The angular factor, w, characterizes the response of a detector to a source at a given sky location and orientation (so that $w = 1$ for an overhead, face-on source). From the distribution of angular factors, w, we assign a distribution of SNRs, $\rho = w\rho_{\rm opt}$, for each source with parameters (m_1, m_2, z). Out of this distribution of SNRs, the fraction that exceed the single-detector

threshold $\rho > 8$, roughly corresponding to a network threshold $\rho > 12$, is taken to be the detection probability $f(z, m_1, m_2)$. This semi-analytic calculation for $f(z, m_1, m_2)$ neglects the effects of non-Gaussian noise, which tends to raise the SNR detection threshold for binaries of very high mass. However, it remains a good approximation for stellar-mass binaries of total masses up to at least $M_{\rm tot} = 100\,M_\odot$ and possibly higher [8, 16].

The expected sensitive redshifted spacetime volume as a function of total mass is shown in Fig. 3.2, calculated for one year of observation ($T = 1$ year) at the O1-O2 LIGO sensitivity. For example, we note that in O1 and O2 the LIGO detectors probed a volume roughly seven times larger for 75–75 $M_\odot$ binaries as compared to 25–25 $M_\odot$ binaries. In particular, since $m_1 = M_{\rm tot}/(1+q)$ for a fixed mass ratio q, we can see from Fig. 3.2 that VT scales approximately as m_1^k with $k \sim 2.2$ for $M_{\rm tot} \lesssim 100\,M_\odot$.

To calculate the sensitivity to a population of BBHs, the relevant quantity is the population-averaged spacetime volume, $\langle VT \rangle$. If we know the distribution of masses across the population of BBHs, $p_{\rm pop}(m_1, m_2)$, assuming negligible spins and a constant comoving merger rate, we can calculate the population-averaged sensitive spacetime volume [Eq. 15 in 8]:

$$\langle VT \rangle = \int\int VT(m_1, m_2) p_{\rm pop}(m_1, m_2)\, dm_1 dm_2, \tag{3.5}$$

where the first integral is over $M_{\rm min} < m_1 < M_{\rm max}$ and the second integral is over $M_{\rm min} < m_2 < \min(m_1,\ M_{\rm max,tot} - m_1)$. $\langle VT \rangle$ relates the specific merger rate, R, to the expected number, Λ, of BBH signals in a given detection period [8]:

$$\Lambda = R\langle VT \rangle. \tag{3.6}$$

The number of BBH detections, n, follows a Poisson process with mean Λ. To explore the existence of a high mass gap in Section 3.5.1, we compare the expected number of low mass BBH signals, $\Lambda_{\rm low} = R\langle VT \rangle_{\rm low}$, to the expected number of high mass BBH signals,

$\Lambda_{\rm high} = R\langle VT\rangle_{\rm high}$, for different power-law populations, where low (high) mass is defined by the primary component mass $m_1 \leq M_{\rm cutoff}$ ($m_1 > M_{\rm cutoff}$). We define:

$$\frac{1}{r} \equiv \frac{\Lambda_{\rm low}}{\Lambda_{\rm high}} = \frac{\langle VT\rangle_{\rm low}}{\langle VT\rangle_{\rm high}}, \tag{3.7}$$

where

$$\begin{aligned}
\langle VT\rangle_{\rm low} &\equiv \int\int_{M_{\rm min}}^{M_{\rm cutoff}} VT(m_1, m_2) p_{\rm pop}(m_1, m_2)\, dm_1 dm_2 \\
\langle VT\rangle_{\rm high} &\equiv \int\int_{M_{\rm cutoff}}^{M_{\rm max}} VT(m_1, m_2) p_{\rm pop}(m_1, m_2)\, dm_1 dm_2.
\end{aligned} \tag{3.8}$$

The integration limits on the m_2 integral in Eq. 3.8 are identical to those in Eq. 3.5 so that the total $\langle VT\rangle = \langle VT\rangle_{\rm low} + \langle VT\rangle_{\rm high}$. We can then compute the probability of detecting $n_{\rm high}$ BBHs with primary component mass $m_1 > M_{\rm cutoff}$, given that we have detected $n_{\rm low}$ BBHs with primary component mass $m_1 < M_{\rm cutoff}$. (We ignore mass measurement uncertainties which may prevent us from definitively assigning a BBH to either the low or high mass class.) This probability is given by:

$$\begin{aligned}
p\left(n_{\rm high} \mid n_{\rm low}\right) &= \int_0^\infty \int_0^\infty p\left(n_{\rm high}, \Lambda_{\rm high}, \Lambda_{\rm low} \mid n_{\rm low}\right) d\Lambda_{\rm high} d\Lambda_{\rm low} \\
&= \int\int p\left(n_{\rm high} \mid \Lambda_{\rm high}\right) p\left(\Lambda_{\rm high} \mid \Lambda_{\rm low}\right) p\left(\Lambda_{\rm low} \mid n_{\rm low}\right) d\Lambda_{\rm high} d\Lambda_{\rm low} \\
&= \int\int p\left(n_{\rm high} \mid \Lambda_{\rm high}\right) \delta\left(\Lambda_{\rm high} - r\Lambda_{\rm low}\right) p\left(\Lambda_{\rm low} \mid n_{\rm low}\right) d\Lambda_{\rm high} d\Lambda_{\rm low} \\
&= \int p\left(n_{\rm high} \mid r\Lambda_{\rm low}\right) p\left(\Lambda_{\rm low} \mid n_{\rm low}\right) d\Lambda_{\rm low} \\
&\propto \int p\left(n_{\rm high} \mid r\Lambda_{\rm low}\right) p\left(n_{\rm low} \mid \Lambda_{\rm low}\right) p_0\left(\Lambda_{\rm low}\right) d\Lambda_{\rm low},
\end{aligned} \tag{3.9}$$

where in the third line we used the definition of r given by Eq. 3.8 and in the last line we used Bayes's theorem. In Eq. 3.9 terms like $p\left(n \mid \Lambda\right)$ denote the Poisson probability of n

with mean Λ. We take the prior $p_0\left(\Lambda_{\rm low}\right)$ to be the Jeffrey's prior:

$$p_0\left(\Lambda_{\rm low}\right) \propto \frac{1}{\sqrt{\Lambda_{\rm low}}}. \tag{3.10}$$

We will return to Eq. 3.9 in Section 3.5.1.

3.4 Fitting the Mass Distribution

Our goal is to jointly infer the shape of the BBH mass distribution along with the lower edge of a potential mass gap, $M_{\rm max}$. We therefore follow [3] in fitting a power-law mass distribution to gravitational-wave BBH mass measurements, but we add the maximum BH mass, $M_{\rm max}$, as a free parameter. We leave the minimum BH mass, $M_{\rm min}$, fixed at $M_{\rm min} = 5\,M_\odot$. Thus, we consider a two-parameter mass distribution:

$$p\left(m_1, m_2 \mid \alpha, M_{\rm max}\right) \propto \frac{m_1^{-\alpha}\,\mathcal{H}(M_{\rm max} - m_1)}{\min(m_1, M_{\rm tot,max} - m_1) - M_{\rm min}}, \tag{3.11}$$

where $\mathcal{H}$ is the Heaviside step function that enforces a cutoff in the distribution at $m_1 = M_{\rm max}$. For consistency with the LIGO definition of a stellar mass BH, we restrict $M_{\rm max} \leq 100\,M_\odot$ throughout. Furthermore, as in the LIGO collaboration's analysis, we enforce $m_1 + m_2 \leq M_{\rm tot,max}$, which provides an additional constraint for $M_{\rm max} > \frac{1}{2}M_{\rm tot,max}$. The LIGO collaboration fixes $M_{\rm tot,max} = 100\,M_\odot$ and $M_{\rm max} = 100\,M_\odot - M_{\rm min}$, as this is the definition of a stellar-mass BBH set by the search. This choice corresponds to one of the following assumptions: either (a) BBHs with total source-frame masses $M_{\rm tot} > 100\,M_\odot$ do not exist as part of the population of stellar-mass BBHs or (b) LIGO is not sensitive to BBHs with total source-frame masses $M_{\rm tot} > 100\,M_\odot$, so we cannot constrain their existence. (In the absence of detections with $M_{\rm tot,max} > 100\,M_\odot$, setting $M_{\rm tot,max} \leq 100\,M_\odot$ in the population model, Eq. 3.11, is equivalent to assuming that the sensitivity vanishes for binaries with $M_{\rm tot} > 100\,M_\odot$.) Assumption (a) may not be well-motivated, as population-synthesis

models that predict stellar BBHs with component masses $M_{\rm max} \sim 50 - 100\,M_\odot$ tend to allow $M_{\rm tot,max} \sim 2M_{\rm max}$ [39, 93]. Assumption (b) can also be questioned, as LIGO is sensitive to BBHs with detector-frame total masses up to at least $600\,M_\odot$ in the IMBH matched-filter search [16]. However, the sensitivity may be lower than expected for very high mass BBHs due to non-stationary instrumental noise [226] and the absence of precessing and higher-order mode template waveforms, which leads to worse matches between signal and template for very high mass BBHs in the matched-filter search [59, 55, 54]. If we had an accurate model of $VT(m_1, m_2)$ across the mass range $5 < m_1, m_2 < 100\,M_\odot$ (by performing a large-scale injection campaign), we could set $M_{\rm tot,max} = 2M_{\rm max} \leq 200\,M_\odot$ in Eq. 3.11. However, because our calculation of VT may be overestimating the sensitivity to binaries with $M_{\rm tot,max} > 100\,M_\odot$, when fitting Eq. 3.11 in the following sections, we repeat the analysis once under the assumption that $M_{\rm tot,max} = \min(2M_{\rm max}, 100\,M_\odot)$ and once assuming that $M_{\rm tot,max} = 2M_{\rm max} \leq 200\,M_\odot$.

To extract the parameters of our assumed mass distribution (Eq. 3.11) from data, we use the same hierarchical Bayesian methods as Appendix D of [3], further explained in [179]. While GW data is noisy and subject to selection effects, both the measurement uncertainties and selection effects are well-quantified. The selection effects refer to the mass-dependent detection efficiency. Under the assumptions of negligible BH spins and a uniform comoving merger rate, the detection efficiency is proportional to the sensitive spacetime volume $VT(m_1, m_2)$ as described in Section 3.3 and [8].

BBH masses are measured using the LALInference parameter-estimation pipeline, which calculates the posterior probability density function (PDF) of all parameters that govern the waveform given the data, d_i, from a BBH detection [245]. For an individual system, measurements of m_1 and m_2 take the form of $\mathcal{O}(10,000)$ posterior samples drawn from the posterior PDF, $p\left(m_1, m_2 \mid d_i\right)$. In the following section, we perform our analysis on published mass measurements from the first four BBHs as well as on simulated BBH measurements. We

use the fact that the one-dimensional PDFs for the source-frame chirp mass and symmetric mass ratio are well-approximated by independent (uncorrelated) Gaussian distributions.

For the first four BBH sources, GW150914, LVT151012, GW151226, GW170104, we approximate the source-frame chirp mass posterior PDF as a Gaussian with a mean and standard deviation given by the median and 90% credible intervals listed in Table 4 of [3] or Table 1 of [11]. In the case that the 90% credible interval is slightly asymmetric about the median, we use the average to estimate the standard deviation. We likewise approximate the posterior PDF of the symmetric mass ratio, $\eta = q/(1+q)^2$, as a Gaussian truncated to the allowed range $[0, 0.25]$, with a mean and standard deviation given by the entry for q in the same tables. Using these approximate chirp mass and symmetric mass ratio distributions, we generate 25,000 posterior samples from the component mass posterior PDFs of each event.

For our set of simulated BBH detections, we generate a set of component masses from an underlying mass distribution. To each BBH system, we assign a redshift from a redshift distribution that is uniform in the merger-frame comoving volume. Given the simulated masses and redshift of each BBH, we randomly generate its single-detector SNR from the antenna power pattern, using the PSD corresponding to the early aLIGO high-sensitivity scenario (as described in Section 3.3). Out of this population, the set of detections are those simulated BBHs with a single-detector SNR satisfying $\rho > 8$. Given the true component masses and the SNR of each mock BBH detection, we produce realistic mass measurements by generating 5,000–10,000 posterior samples for the component masses following the prescription in Eq. 1 of [175]. Given true values for the chirp mass, $\mathcal{M}^T$, symmetric mass ratio, η^T, and SNR, ρ^T, we draw chirp mass posterior samples from a Gaussian distribution centered at $\bar{\mathcal{M}}$ with standard deviation $\sigma_{\mathcal{M}}$ and symmetric mass ratio posterior samples from a Gaussian distribution centered at $\bar{\eta}$ with standard deviation σ_η. We only keep posterior samples with

$0.01 \lesssim \eta \lesssim 0.25$. The variables $\bar{\mathcal{M}}$ and $\bar{\eta}$ are drawn from Gaussian distributions:

$$\begin{aligned} \bar{\mathcal{M}} &\sim N(\mathcal{M}^T, \sigma_{\mathcal{M}}), \\ \bar{\eta} &\sim N(\eta^T, \sigma_\eta), \end{aligned} \tag{3.12}$$

where $\sigma_{\mathcal{M}}$, σ_η scale inversely with the SNR, and are given in [175].

Once we have samples from the posterior PDF, $p\left(m_1, m_2 \mid d_i\right)$, for each event (both real and simulated) and we have calculated the detection efficiency, $P_{\rm det}(m_1, m_2) \propto VT(m_1, m_2)$, we follow Appendix D in [3] to fit Eq. 3.11. The likelihood for a single BBH detection given the parameters of the mass distribution, α and $M_{\rm max}$, is given by:

$$\begin{aligned} p\left(d_i \mid \alpha, M_{\rm max}\right) &\propto \frac{\int p\left(d_i \mid m_1, m_2\right) p\left(m_1, m_2 \mid \alpha, M_{\rm max}\right) dm_1 dm_2}{\beta\left(\alpha, M_{\rm max}\right)} \\ &\propto \frac{\left\langle p\left(m_1, m_2 \mid \alpha, M_{\rm max}\right)\right\rangle}{\beta\left(\alpha, M_{\rm max}\right)}, \end{aligned} \tag{3.13}$$

where $\langle \ldots \rangle$ denotes an average over the (m_1, m_2) posterior samples. This is valid because for each event, $p\left(d_i \mid m_1, m_2\right) \propto p\left(m_1, m_2 \mid d_i\right)$, as the prior on $m_1,\ m_2$ is taken to be flat. Therefore, we can calculate the integral in the first line of Eq. 3.13 by taking the average of $p(m_1, m_2 \mid \alpha, M_{\rm max})$ over the mass posterior samples. Meanwhile, $\beta(\alpha, M_{\rm max})$ is defined as:

$$\beta(\alpha, M_{\rm max}) \equiv \int p\left(m_1, m_2 \mid \alpha, M_{\rm max}\right) VT(m_1, m_2)\, dm_1 dm_2. \tag{3.14}$$

The likelihood for the data across all events $\mathbf{d} = \{d_i\}$ is the product of the individual event likelihoods given by Eq. 3.13.

Furthermore, if we fix $M_{\rm max}$ and assume a prior $p_0\left(\alpha \mid M_{\rm max}\right)$, we can calculate the Bayesian evidence in favor of a given $M_{\rm max}$:

$$p\left(d_i \mid M_{\rm max}\right) = \int p\left(d_i \mid \alpha, M_{\rm max}\right) p_0\left(\alpha \mid M_{\rm max}\right) d\alpha. \tag{3.15}$$

We can then calculate the Bayes factor between two power-law models that differ in their choice of $M_{\rm max}$. Recall that the default LIGO analysis fixes $M_{\rm max} = 100\,M_\odot - M_{\rm min}$. For a sample of N detected BBHs (assumed to be independent), the cumulative Bayes factor $K(M_{\rm max}, 100\,M_\odot)$ between a power-law model that fixes $M_{\rm max} = M$ and one that fixes $M_{\rm max} = 100\,M_\odot$ is a product of the single-event evidence ratios:

$$K(M, 100\,M_\odot) = \prod_{i=1}^{N} \frac{p\left(d_i \mid M_{\rm max} = M\right)}{p\left(d_i \mid M_{\rm max} = 100\,M_\odot\right)}. \tag{3.16}$$

We calculate the cumulative Bayes factor $K(M, 100\,M_\odot)$ in Section 3.5.2.

3.5 Results

3.5.1 Non-detection of heavy BBHs

We first give an example of how the detection of only four BBHs with primary masses $m_1 \leq 40$–$50\,M_\odot$ is inconsistent with certain (possibly correct) power-law mass distributions unless a mass gap is imposed. For a given mass distribution, we can use Eq. 3.9 to calculate the probability of not detecting any BBHs above a certain mass, $n_{\rm high} = 0$, given that we have detected $n_{\rm low}$ BBHs below the cutoff mass. To do this, we must first compute the ratio $\langle VT\rangle_{\rm low}/\langle VT\rangle_{\rm high}$ as defined in Eq. 3.7, and then apply Eq. 3.9. The results of this calculation for $p(n_{\rm high} = 0 \mid n_{\rm low} = 4)$ and $p(n_{\rm high} = 0 \mid n_{\rm low} = 10)$ are displayed in Fig. 3.5.1. We show the results for two choices of cutoff mass: $M_{\rm cutoff} = 41\,M_\odot$ (green curve) is motivated by the 95% credible upper limit on the primary mass of GW150914, the heaviest BBH detected, and $M_{\rm cutoff} = 50\,M_\odot$ (blue and orange curves) is motivated by PPISN and PISN supernova models, which predict a mass gap starting at 40–50 $M_\odot$ (depending also on details of binary evolution) [38, 254]. We also vary the maximum total mass, $M_{\rm tot,max}$, of the "high mass" population between $100\,M_\odot$ (blue and green curves), which is currently

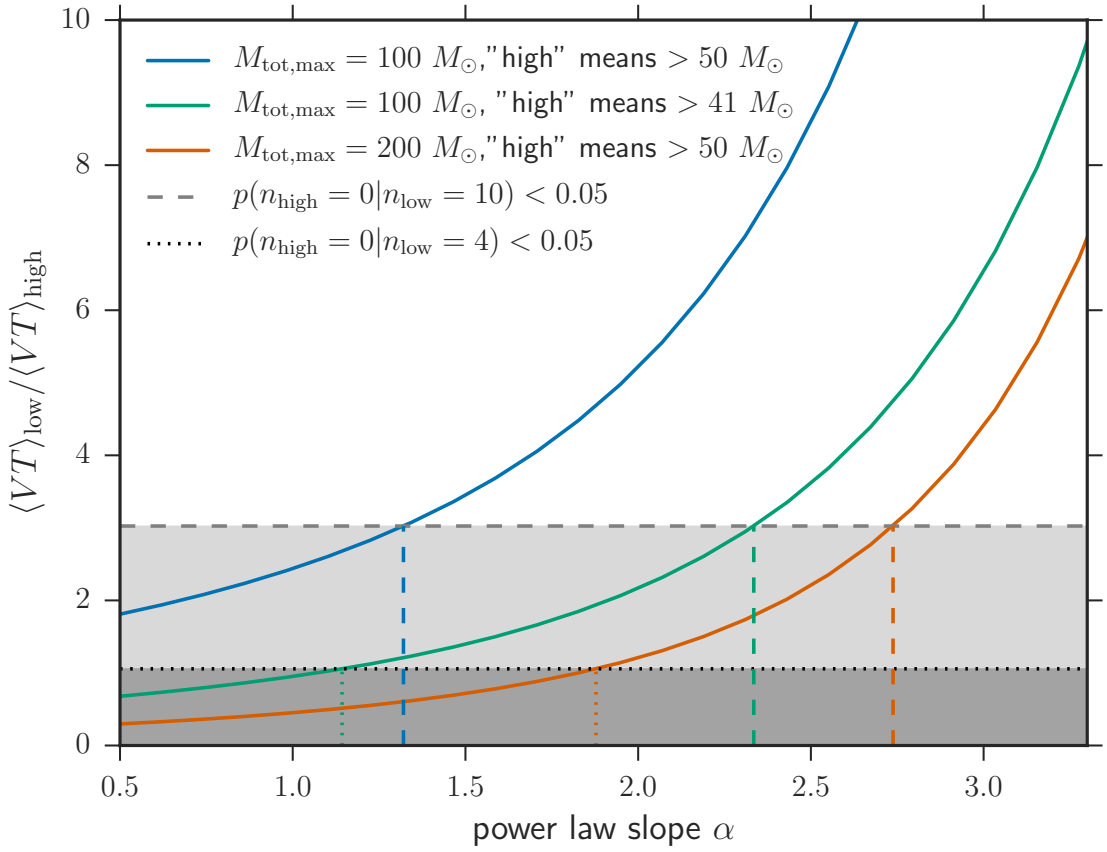

Figure 3.2: **Ratio $\langle VT\rangle_{\rm low}/\langle VT\rangle_{\rm high}$ of the expected number of"low" mass to "high" mass BBH detections** from an underlying population with power law slope α (solid curves). The blue and orange curves define low mass BBHs to have $m_1 < 50\,M_\odot$, while the green curve defines low mass as $m_1 < 40\,M_\odot$. The blue and green curves conservatively assume that the mass distribution and detector sensitivity extend only up to a total BBH mass of $100\,M_\odot$, while the orange curve assumes that the BBH population and sensitivity extend up to total masses of $200\,M_\odot$. The dashed (dotted) horizontal line corresponds to values of the ratio $\langle VT\rangle_{\rm low}/\langle VT\rangle_{\rm high}$ for which the probability of detecting no high mass BBHs and ten (four) low mass BBHs is less than 5%. VT ratios below this line correspond to values of α that lie to the left of the vertical colored dashed lines. With enough low mass BBH detections, shallow power law slopes (small positive values of α) become inconsistent with the existence of high mass BBHs.

the maximum total mass that the aLIGO search includes in the definition of a stellar mass BBH, to $200\,M_\odot$ (orange curve). We note that a power law with slope $\alpha = 1$ is the "flat in log" population that the LIGO-Virgo collaboration uses to compute the lower limits on the BBH merger rate, and we calculate that unless a mass gap is imposed, detecting four BBHs with primary masses $m_1 < 41\,M_\odot$ is inconsistent with this population at the 96% level if we restrict $M_{\rm tot,max} = 100\,M_\odot$, or at the $> 99.9\%$ level if we assume that the BBH population and the detectors' sensitivity extends up to $M_{\rm tot,max} = 200\,M_\odot$. Furthermore, unless a mass gap is imposed, there is already some tension (inconsistency at the 93% level) with the $\alpha = 2.35$ population that LIGO-Virgo uses to compute the upper rate limits if we assume $M_{\rm tot,max} = 200\,M_\odot$. If the BBH mass distribution has an upper cutoff at 40–50 $M_\odot$, the inferred merger rates calculated without assuming the cutoff would be 1.4–2.1 times higher for the "flat in log" population and 1.1–1.4 times higher for the $\alpha = 2.35$ population.

3.5.2 Bayesian evidence in favor of mass gap

We have seen that assuming a single power-law mass distribution over the entire mass range 5–100 $M_\odot$ can rule out shallow power law slopes in the absence of detections with component masses $m_1 > 40$–$50\,M_\odot$. The absence of high mass detections will continue to push the inferred power law slope to steeper values unless we allow for an upper mass gap in the analysis. To study this point further, we simulate mock BBH mass measurements from a power-law population with slope $\alpha = 2.35$ and an upper mass cutoff at $M_{\rm max} = 41\,M_\odot$) (Eq. 3.11, but we follow the canonical analysis used by the LVC (see Eq. 3.3) and fix $M_{\rm max} = 95\,M_\odot$ and $M_{\rm tot,max} = 100\,M_\odot$ when inferring the power law slope. While the bias on the inferred slope α may be small with $\mathcal{O}(10)$ detections, with $\mathcal{O}(100)$ detections, the canonical analysis will rule out the correct power law slope (see Fig. 3.5.2). (Although for hundreds of detections, a non-parametric fit to the mass distribution should be considered.) If we follow the canonical LIGO-Virgo analysis but set $M_{\rm max} = 100\,M_\odot$ and $M_{\rm tot,max} = 200\,M_\odot$ rather

than $M_{\rm tot,max} = 100\,M_\odot$, the presence of a mass gap will bias the power law inference even more significantly, as the true population has $M_{\rm tot,max} = 82\,M_\odot$. These results show that failing to account for an upper mass gap may lead to incorrect conclusions about the low mass distribution. While we demonstrated this for an assumed power law model, this caveat applies to any parametrized fit to the BBH mass distribution.

With the first four LIGO BBH detections, varying $M_{\rm max}$ does not drastically bias the inference on α when fitting Eq. 3.3 (see the solid lines in Fig. 3.5.2). However, we can distinguish the model favored by the data by calculating the cumulative Bayes factor. Following Eqs. 3.15–3.16, we calculate this factor between two single-parameter power law models with different fixed values of $M_{\rm max}$. We choose to compare two cutoff values, $41\,M_\odot$ (the 95% upper limit on the heaviest component BH detected) and $100\,M_\odot$, and take the prior $p_0\left(\alpha \mid M_{\rm max}\right)$ in Eq. 3.15 to be a top hat over the wide range $-2 < \alpha < 7$.

For the first four BBH detections, $K(41\,M_\odot, 100\,M_\odot) = 13$ if we assume the detectable BBH population only extends to $M_{\rm tot,max} = 100\,M_\odot$ (so that $M_{\rm max} = 100\,M_\odot$ is really $M_{\rm max} = 95\,M_\odot$). If we instead assume that the $M_{\rm max} = 100\,M_\odot$ population is fully detectable up to total binary masses of $M_{\rm tot,max} = 200\,M_\odot$, the Bayes factor increases to $K(41\,M_\odot, 100\,M_\odot) = 90$, suggesting that there is already strong support for an upper mass cutoff at $M_{\rm max} \sim 40\,M_\odot$ over a cutoff at $M_{\rm max} \sim 100\,M_\odot$ within the assumed power law model [145]. The Bayes factor also depends on the choice of prior on α. We choose to be relatively uninformative in our prior, excluding only very steeply declining mass distributions ($\alpha > 7$) and allowing for moderately upward sloping mass distributions ($-2 < \alpha < 0$), but it is clear from Fig. 3.5.1 that a prior that favors large positive values of α (steeply declining power law slopes) will lower the evidence in favor of a mass cutoff $M_{\rm max} < 100\,M_\odot$, while placing greater prior support on low values of α (shallow or downward sloping power laws) will raise the evidence in favor of a mass cutoff. If we enforce $\alpha > 0$ in the prior in order to agree with other astrophysical mass distributions, the Bayes factors change to

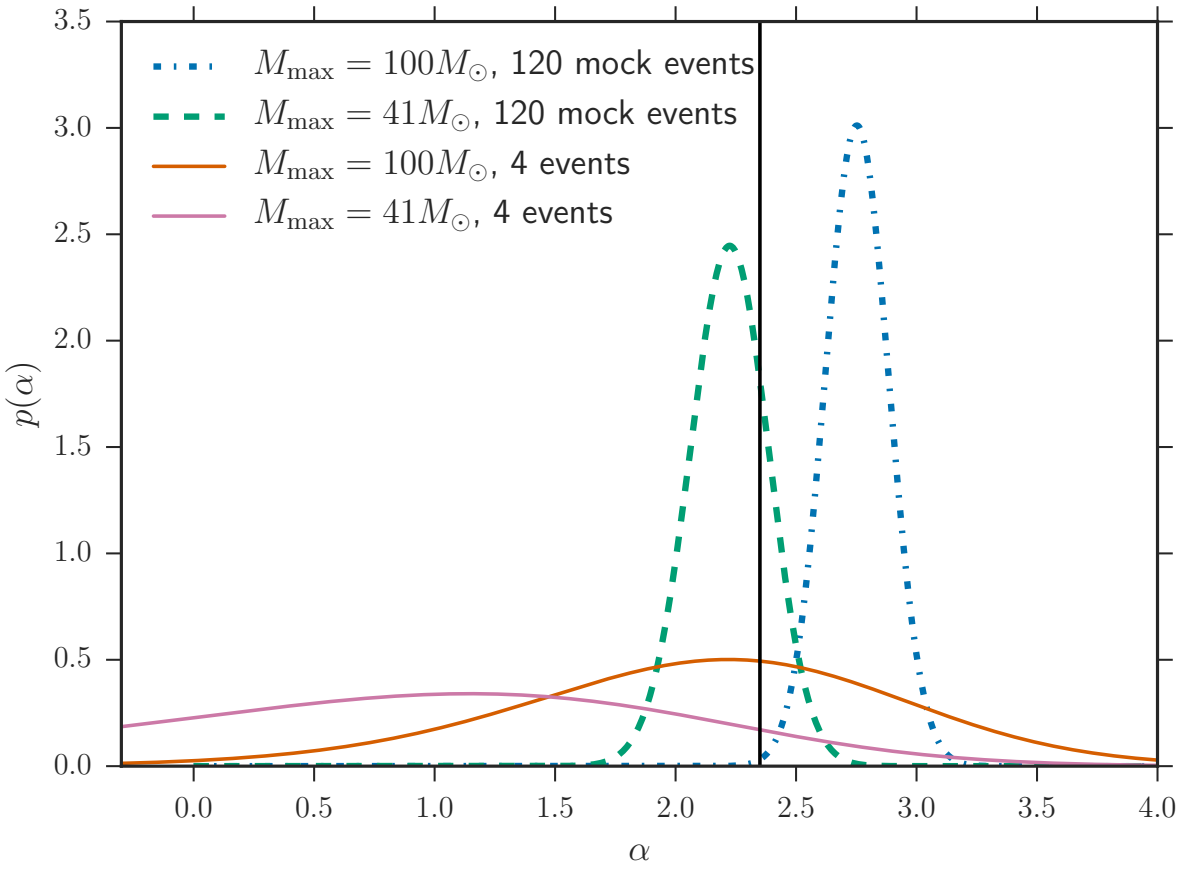

Figure 3.3: **Inferred likelihood for the power law slope of the mass distribution**, α, calculated for 120 mock observations from a $M_{\max} = 41\,M_\odot$, $\alpha = 2.35$ population (dashed and dotted curves) and the first four BBH detections (solid curves). The blue and orange curves correspond to the canonical LVC analysis in which the maximum mass of the BBH mass distribution is set to $M_{\max} = 95\,M_\odot$, while the pink and green curves correspond to a fixed maximum mass at $M_{\max} = 41\,M_\odot$. Neglecting to account for a high mass cutoff biases the power law inference to steep slopes. The solid black line at $\alpha = 2.35$ is the true slope of the simulated population, but gets ruled out by the canonical analysis.

$K(41\,M_\odot, 100\,M_\odot) = 5$ if we restrict $M_{\rm tot,max} \leq 100\,M_\odot$ or $K(41\,M_\odot, 100\,M_\odot) = 21$ if we allow $M_{\rm tot,max} \leq 200\,M_\odot$.

We anticipate that a set of ten BBH detections with primary component masses $m_1 \leq 41\,M_\odot$ will yield a Bayes factor $K(41\,M_\odot, 100\,M_\odot) > 150$, providing very strong evidence for an upper mass gap. We assume that the underlying BBH population (and aLIGO's sensitivity) extends to total masses of $2M_{\rm max}$ in either case, so $M_{\rm tot,max} \leq 200\,M_\odot$. We take a flat prior on α in the range $[-2, 7]$. With 191 events from a simulated $\alpha = 2.35$, $M_{\rm max} = 41\,M_\odot$ population, the single-event evidence ratios range from $K(41\,M_\odot, 100\,M_\odot) = 1.0$ to $K(41\,M_\odot, 100\,M_\odot) = 16.9$, with a median of $K(41\,M_\odot, 100\,M_\odot) = 2.6$. With a subset of ten BBH detections from this population, we get $K(41\,M_\odot, 100\,M_\odot) > 150$ in more than 99% of cases. If we detect ten BBHs with primary component masses $m_1 \leq 50\,M_\odot$, we likewise expect very strong evidence for a mass cutoff, with $K(50\,M_\odot, 100\,M_\odot) > 150$ more than 95% of the time. The Bayes factor only compares two values of the mass cutoff; we fit for the value of $M_{\rm max}$ favored by a given set of detections in the following subsection 3.5.3.

3.5.3 Joint power law–maximum mass fit

In this section we fit the two-parameter mass distribution of Eq. 3.11. We calculate the likelihood $p(\mathbf{d} \mid \alpha, M_{\rm max})$ as the product of individual event likelihoods in Eq. 3.13. We take flat priors on α and $M_{\rm max}$, with $-2 \leq \alpha \leq 7$ as before and $M_{\rm max} \leq 100\,M_\odot$. The minimum allowed value of $M_{\rm max}$ for a given set of detections is set by the lower mass bound of the heaviest detected component BH. For simplicity, we take the lower mass bound to be the minimum posterior sample. The upper bound $M_{\rm max} \leq 100\,M_\odot$ is motivated by the LIGO stellar mass BBH search as well as by population synthesis studies, which usually predict that the BH mass distribution would extend to 80–130 $M_\odot$ were it not for a pair-instability mass gap [93, 38, 230]. We calculate the likelihood function on a 500×100 grid of $(\alpha, M_{\rm max})$ values in the allowed prior range, and verify that increasing the resolution of the

$(\alpha, M_{\rm max})$ grid does not change our results. In fact, the resolution in the $M_{\rm max}$ dimension is limited by the finite number of posterior samples that are used to represent the component mass posterior PDFs for each event. To reduce these artificial discontinuities in the $M_{\rm max}$ dimension of the likelihood evaluation, we apply a two-dimensional smoothing spline before displaying the results.

The results of the joint power law–maximum mass analysis for the set of four detected BBHs is shown in Fig. 3.4. We compute the joint likelihood twice: once fixing $M_{\rm tot,max} = 100\,M_\odot$, so that the population of stellar mass BBHs is restricted to total masses $M_{\rm tot} \leq 100\,M_\odot$ regardless of $M_{\rm max}$ (top right panel) and once fixing $M_{\rm tot,max} = 200\,M_\odot$, so that the maximum total mass of the population is allowed to extend to $M_{\rm tot,max} = 2M_{\rm max}$. We calculate the marginal posterior PDFs of α and $M_{\rm max}$ (top left and bottom right panels) under the assumption of a uniform prior on α in the range $[-2, 7]$ and a uniform prior on $M_{\rm max}$ in the range $[29\,M_\odot, 100\,M_\odot]$ ($29\,M_\odot$ is the minimum posterior sample we generated for the primary component mass of GW150914).

It is clear that properly accounting for our uncertainty on $M_{\rm max}$ when fitting the power-law mass distribution increases the support for shallow power law slopes which would otherwise be ruled out under the assumption that the mass distribution extends continuously to $\sim 100\,M_\odot$. Allowing for freedom in $M_{\rm max}$ shifts the preferred values of α to shallower slopes, even allowing for negative α, as compared to the canonical analysis that fixes $M_{\rm max} = 95\,M_\odot$ (orange dot-dashed curve in Fig. 3.4). Furthermore, the first four BBH detections already start to constrain $M_{\rm max}$. The marginal posterior PDF $p(M_{\rm max})$ peaks strongly at $M_{\rm max} \sim 40$, and the 95% upper limits on the inferred $p(M_{\rm max})$ are $76.6\,M_\odot$ if assuming $M_{\rm tot,max} = 200\,M_\odot$ (or $90.7\,M_\odot$ if we conservatively assume $M_{\rm tot,max} = 100\,M_\odot$). Taking $M_{\rm tot,max} = 200\,M_\odot$ rather than $100\,M_\odot$ allows the detectable BBH population to extend to $2M_{\rm max}$, thereby increasing the expected sensitivity to BBHs with primary component masses $M_{\rm max} > 50\,M_\odot$. Thus, the non-detection of heavy BBHs yields tighter

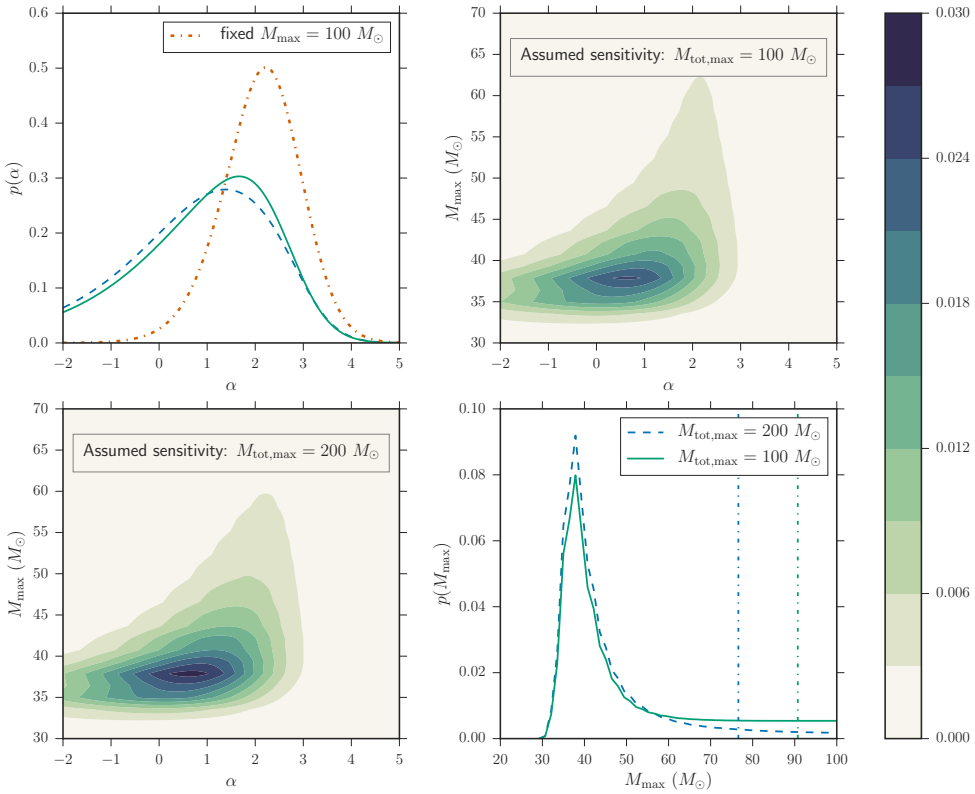

Figure 3.4: **Joint fits for α, $M_{\rm max}$ from the first four LIGO detections.** Lower left panel: The posterior PDF $p(\alpha, M_{\rm max})$ under the conservative assumption that $M_{\rm tot,max} = 100\,M_\odot$. Upper right panel: The posterior PDF $p(\alpha, M_{\rm max})$ assuming full matched-filter sensitivity up to $M_{\rm tot,max} = 200\,M_\odot$. Upper left panel: The marginal posterior PDF for α under each assumption of $M_{\rm tot,max}$ (green solid and dashed blue curves). The orange dash-dotted curve shows the results of the canonical analysis (Eq. 3.3) in which $M_{\rm max}$ is fixed at $M_{\rm max} = 95\,M_\odot$ and $M_{\rm tot,max} = 100\,M_\odot$. Lower right panel: The marginal posterior PDF for $M_{\rm max}$ under each assumption of $M_{\rm tot,max}$. The vertical dotted lines denote 95% credible intervals. Throughout, we take a uniform prior on α in the range $-2 \leq \alpha \leq 7$ and on $M_{\rm max}$ in the range $M_{\rm max} < 100\,M_\odot$, as described in the text.

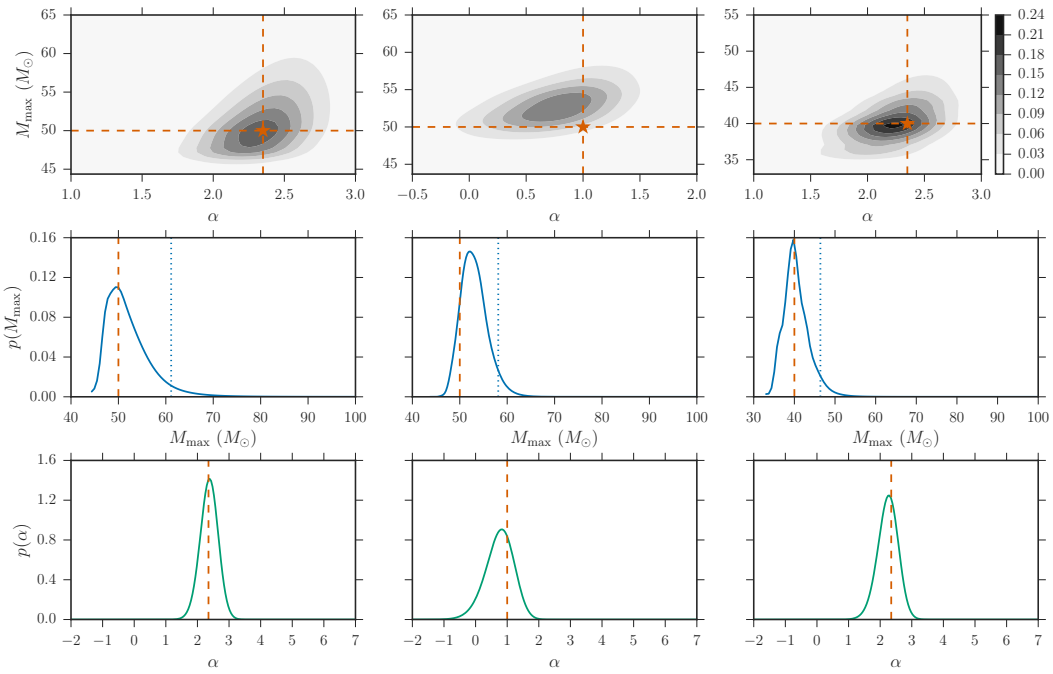

Figure 3.5: **Joint fits for (α, $M_{\rm max}$) using 40 simulated BBH detections from 3 populations**, assuming that $M_{\rm tot,max} = 2M_{\rm max}$ and that LIGO/Virgo is sensitive up to $M_{\rm tot,max} = 200\,M_\odot$. Each column represents a different simulated population where the true α, $M_{\rm max}$ values are shown by the orange star. Left column: $\alpha = 2.35$, $M_{\rm max} = 50\,M_\odot$; middle column: $\alpha = 1$, $M_{\rm max} = 50\,M_\odot$; right column: $\alpha = 2.35$, $M_{\rm max} = 40\,M_\odot$. The top row shows the posterior PDF $p(\alpha, M_{\rm max})$ as recovered from 40 events, the second row shows the marginal PDF of $M_{\rm max}$, and the bottom row shows the marginal PDF of α for each simulated population.

constraints on the inferred $M_{\rm max}$ when we assume $M_{\rm tot,max} = 200\,M_\odot$, but the peak of the $M_{\rm max}$ distribution remains unchanged.

To explore the impact of future detections on the inferred mass distribution, we repeat this analysis for three simulated BBH populations, two with a power law slope $\alpha = 2.35$ and one with a power law slope $\alpha = 1$. One of the $\alpha = 2.35$ populations, as well as the $\alpha = 1$ population, has a mass gap starting at $M_{\rm max} = 50\,M_\odot$, while the other $\alpha = 2.35$ population has a mass gap starting at $M_{\rm max} = 40\,M_\odot$.

The results of this calculation, assuming sensitivity up to $M_{\rm tot,max} = 200\,M_\odot$, are shown

in Fig. 3.5, where each column corresponds to 40 detections from one of the three simulated populations. We see that 40 detections yield strong constraints on both the slope and maximum mass of the population. If the true population has a cutoff at $M_{\rm max} = 40\,M_\odot$ (right column in Fig. 3.5) rather than $50\,M_\odot$, we get tighter constraints on $M_{\rm max}$. Similarly, we expect better constraints on the maximum mass for shallower mass distributions, as the non-detection of heavy BHs is more striking for shallow mass distributions (see Fig. 3.5.1). The maximum mass is indeed better constrained for the population with power law slope $\alpha = 1$ (middle column) than for the population with the same mass cutoff $M_{\rm max} = 50\,M_\odot$ but steeper power law slope $\alpha = 2.35$ (left column). As the first four LIGO detections currently favor $M_{\rm max} < 50\,M_\odot$ and moderately shallow power law slopes $\alpha < 3$ (Fig. 3.4), we expect that fewer than 40 detections will strongly constrain $M_{\rm max}$.

3.6 Discussion

3.6.1 Effect of Redshift Evolution

We have assumed that the merger rate, as measured in the source-frame, is uniform in comoving volume (Eq. 3.4). In reality, the merger rate per comoving volume is expected to increase with redshift until $z \sim 2$ [see, for example, Extended Data Fig. 4 in 39]. This would mean that we have underestimated the VT factors for high mass systems, because high mass systems are detectable at higher redshifts. Thus, we have also underestimated the terms $\langle VT\rangle_{\rm low}/\langle VT\rangle_{\rm high}$ displayed as a function of power law slope α in Fig. 3.5.1, and the tension between certain power law slopes and the non-detection of heavy BHs is in fact greater than we predicted. If we assumed a steeper redshift evolution of the merger rate, fewer detections would resolve the mass gap at high confidence.

3.6.2 Distribution of mass ratios

In fitting the mass distribution of BBHs (Eq. 3.11), we assumed that the distribution of mass ratios, q, is uniform in the allowed range $M_{\rm min}/m_1 < q < \min(M_{\rm tot,max}/m_1 - 1, 1)$. In particular, we assumed that for a given primary component mass, m_1, the marginal distribution of $m_2 = qm_1$ is given by Eq. 3.2. However, many BBH formation models predict a preference for equal-mass mergers [86, 211]. To explore the effects of our assumed mass ratio distribution, we generalize Eq. 3.2:

$$p(m_2 \mid m_1) \propto \frac{m_2^k}{\min(m_1, M_{\rm tot,max} - m_1)^{k+1} - M_{\rm min}^{k+1}}, \tag{3.17}$$

so that $k = 0$ reduces to Eq. 3.2 while $k > 0$ favors more equal mass ratios. We find that the choice of $k \geq 0$ does not noticeably impact our results, and we recover consistent posteriors on $(\alpha, M_{\rm max})$ if we fix $k = 6$ rather than $k = 0$. However, we note that there is currently no evidence that the distribution of mass ratios, $p(q \mid m_1)$, deviates from the assumed uniform distribution. Although all of the events so far are consistent with mass ratios close to unity, this is not surprising given the selection effects that favor more equal-mass systems. For a fixed primary mass, m_1, assuming full matched-filter sensitivity, we would expect five detections with $q > 0.5$ for every detection with $q < 0.5$, and two detections with $q > 0.7$ for every detection with $q < 0.7$, even if we take q to be uniform in the range $[0, 1]$ rather than $[M_{\rm min}/m_1, 1]$. We can explicitly check if the data favors $k > 0$ if we incorporate Eq. 3.17 into the power law mass distribution, so that Eq. 3.11 becomes:

$$p\left(m_1, m_2 \mid \alpha, M_{\rm max}\right) \propto \frac{m_1^{-\alpha} m_2^k \, \mathcal{H}(M_{\rm max} - m_1)}{\min(m_1, M_{\rm tot,max} - m_1)^{k+1} - M_{\rm min}^{k+1}}. \tag{3.18}$$

We fit the above Eq. 3.18 for k, marginalizing over α and $M_{\rm max}$, and find that, for the first four LIGO detections, the likelihood $p(\mathbf{d} \mid k)$ peaks mildly at $k = 0$, but is very broad. Thus, the first four BBHs mildly favor a uniform distribution of mass ratios. Future detections will

continue to test this assumption.

3.6.3 Extending to non-power-law mass distributions

Although a power law provides a good fit to the mass distribution of massive stars, there are theoretical indications that the masses of BHs in merging binaries may diverge from a power-law distribution. For example, supernova theory suggests that there is a nonlinear relationship between the initial zero-age main sequence mass of star and its resulting BH [39, 230]. In fact, PPISN and PISN are associated with significant mass loss and may cause a deviation in the BH mass distribution at masses $> 30\,M_\odot$. Additionally, several models predict that a mass-dependent merger efficiency causes the mass distribution for merging binaries to differ significantly from the BH mass function [198]. While we have focused solely on power law fits to the mass distribution, an increased sample of BBH detections will allow us to explore more complicated parametric and non-parametric models and select a model for the mass distribution that best fits the data. Regardless of the model, it is straightforward to include a free parameter (in our case, $M_{\rm max}$) that fits for the bottom edge of the upper mass gap.

3.6.4 Are there BBHs beyond the gap?

So far we have restricted our attention to the bottom edge of the upper mass gap, but LIGO is also probing the upper edge of the mass gap in the IMBH search, with results from the first observing run presented in [16]. It is theoretically unclear whether BHs exist on the other side of the mass gap (predicted at $\sim 135\,M_\odot$), as the frequency of sufficiently high mass stars is unknown [39]. Before accounting for PPISN or PISN, previous population-synthesis predictions placed the maximum BH mass at 80–$135\,M_\odot$ for zero age main sequence masses $M_{\rm ZAMS} < 150\,M_\odot$ [38, 93, 230]. However, stars with $M_{\rm ZAMS} \gtrsim 200\,M_\odot$ in a sufficiently low-metallicity environment ($Z \sim 0.07 Z_\odot$) are expected to directly collapse to BHs with

masses ~ 120–$280\,M_\odot$ [231]. We find that LIGO is approximately five times more sensitive to a population of equal mass BBHs just above the mass gap (with total masses in the range 270–$300\,M_\odot$) than to equal mass BBHs with total masses in the range 10–$65\,M_\odot$. In the unlikely scenario that a power law continues unbroken over the entire mass range $10 \leq M_{\rm tot} \leq 300\,M_\odot$, we could constrain the existence of BBHs above the mass gap by extrapolating the power-law fit from the mass distribution below the gap. If we take a power law with slope $\alpha = 1$, the expected number of detected BBHs below the gap ($5 \lesssim m_1 \lesssim 40$) is ~ 1.76 times greater than the expected number of detected binaries directly above the gap ($135 \leq m_1 \leq 150\,M_\odot$). This means that within ~ 10 BBH detections with $5 \lesssim m_1 \lesssim 40$, the non-detection of BBHs above the gap would imply that the power-law extrapolation with $\alpha = 1$ breaks down or that BBHs above the gap do not exist. However, if we extrapolate a power law with slope $\alpha = 2.35$ across this mass range, the expected number of detections below the gap is ~ 22.9 times the expected number of detections directly above the gap, so it would take $\gtrsim 60$ BBH detections to invalidate the power law extrapolation to the other side of the gap. The existing sample of BBHs is too small to place interesting constraints on the existence of systems beyond the gap.

3.7 Conclusion

We have shown that given LIGO's extremely high sensitivity to BBHs with component masses $40 \leq m_1 \leq 100\,M_\odot$, it is statistically significant that the first four detections have been less massive than $40\,M_\odot$. We present a two-parameter model for the BBH mass distribution, consisting of a power law with slope α and a cutoff at $M_{\rm max}$, and find that the first four detections already provide evidence for a cutoff to the mass distribution at $M_{\rm max} \sim 40\,M_\odot$. This cutoff may be the lower edge of a PPISN/ PISN upper mass gap. Furthermore, LIGO-Virgo have recently announced two more BBH detections (GW170608 and GW170814), both of which are less massive than $40\,M_\odot$ and only strengthen our conclusions

[13, 12]. We find that within $\mathcal{O}(10)$ BBH detections, the location of the bottom edge of the upper mass gap will be significantly constrained. Our model assumes that all BBHs belong to a single population described by the same power law, so that the detection of a binary with mass in the mass "gap" would reset the lower edge of the gap beyond the mass of the newly-detected binary. However, we expect to quickly converge on the true maximum mass of the population within $\lesssim 40$ detections. At this point, the detection of a binary in the mass gap will be statistically inconsistent with this single population, and may indicate a subpopulation of BBHs that did not form directly from stellar-collapse (e.g. primordial BHs or BHs formed through previous mergers). The BBH spin distribution will provide further constraints on the existence of these subpopulations and will allow us to measure the fraction of BBHs that formed through previous mergers [106, 120].

CHAPTER 4

DOES THE BLACK HOLE MERGER RATE EVOLVE WITH REDSHIFT?

The previous two chapters discussed how we can learn about the origins of binary black holes from their masses and spins. This chapter looks at a third property of the binary black hole population: how many black hole mergers are there, and does the merger rate change over the history of the universe? Using the first six binary black hole mergers detected by LIGO and Virgo, we fit the leading order redshift evolution of the merger rate, and found that the rate was consistent with a non-evolving merger rate, as well as astrophysically-reasonable deviations. We were able to rule out extreme deviations that would point to, for example, strong lensing of the gravitational-wave signals or deviations from general relativity.

The redshift evolution of the merger rate contains information regarding the formation times of binary black hole progenitors, and the time delays between formation and redshift. For example, comparing the merger rate to the evolution of stars, galaxies, and chemical enrichment over cosmic time will shed insight into the environments and processes that affect black hole formation and merger. In this chapter, we find that with roughly 500 detections by LIGO at design sensitivity, we will be able to meaningfully distnguish between different formation channels for binary black holes by measuring the local redshift evolution. For example, we will be able to measure whether the merger rate increases faster or slower than the star formation rate locally. Building upon the work in this chapter, we recently collaborated with Tom Callister to show that combining this measurement with information from the stochastic gravitational-wave background – which consists of all of the quiet, unresolved mergers in the universe – will considerably improve this measurement, allowing us to constrain the merger rate out to redshifts beyond $z > 2$ and measure the location of a peak in the rate [56].

This chapter was carried out in collaboration with Will Farr and Daniel Holz and pub-

lished as [109]. This chapter contains the analysis on the first six binary black hole detections. I later repeated this analysis on the full O1 and O2 sample consisting of ten binary black holes in [19]. Analyzing the full dataset, we found a preference for merger rates that increased with increasing redshift, ruling out merger rates that declined with redshift at the 93% level.

MF was supported by the NSF Graduate Research Fellowship Program under grant DGE-1746045. MF and DEH were supported by NSF grant PHY-1708081. They were also supported by the Kavli Institute for Cosmological Physics at the University of Chicago through NSF grant PHY-1125897 and an endowment from the Kavli Foundation. DEH also gratefully acknowledges support from the Marion and Stuart Rice Award. WMF is supported in part by the STFC.

4.1 Abstract

We explore the ability of gravitational-wave detectors to extract the redshift distribution of binary black hole (BBH) mergers. The evolution of the merger rate across redshifts $0 < z \lesssim 1$ is directly tied to the formation and evolutionary processes, providing insight regarding the progenitor formation rate together with the distribution of time delays between formation and merger. Because the limiting distance to which BBHs are detected depends on the masses of the binary, the redshift distribution of detected binaries depends on their underlying mass distribution. We therefore consider the mass and redshift distributions simultaneously, and fit the merger rate density, $dN/dm_1\,dm_2\,dz$. Our constraints on the mass distribution agree with previously published results, including evidence for an upper mass cutoff at $\sim 40\ M_\odot$. Additionally, we show that the current set of six BBH detections are consistent with a merger rate density that is uniform in comoving volume. Although our constraints on the redshift distribution are not yet tight enough to distinguish between BBH formation channels, we show that it will be possible to distinguish between different

astrophysically motivated models of the merger rate evolution with $\sim$ 100–300 LIGO-Virgo detections (to be expected within 2–5 years). Specifically, we will be able to infer whether the formation rate peaks at higher or lower redshifts than the star formation rate, or the typical time delay between formation and merger. Meanwhile, with $\sim$ 100 detections, the inferred redshift distribution will place constraints on more exotic scenarios such as modified gravity.

4.2 Introduction

The redshift dependence of the binary black hole (BBH) merger rate carries information about the processes by which BBHs evolve and merge, including the environments in which they form, the star formation rate (SFR), and the time-delay distribution. By measuring the luminosity distance[1] to detected sources, the current generation of ground-based gravitational-wave (GW) detectors will be able to measure the redshift distribution of BBH mergers up to redshifts $z \sim 1$ [2, 7]. The inferred redshift distribution will provide important clues regarding the BBH formation channel. For example, in the classical isolated binary evolution channel, the redshift evolution follows the SFR convolved with a distribution of time delays between formation and merger [91, 2, 39]. Meanwhile, in the dynamical formation channel, the evolution of BBH mergers is tied to the evolution of globular clusters [65, 213]. If BBHs are primordial, they are expected to largely follow the dark matter distribution [181, 157]. Furthermore, several exotic scenarios, such as gravitational leakage [92, 80, 203] or a significant population of strongly lensed BBH systems [51, 228], would leave an imprint on the inferred redshift distribution.

In this work we consider the BBH merger rate density as a function of the component masses, m_1 and m_2, and redshift, z. The mass and redshift distributions must be fit si-

1. Throughout we fix the cosmological parameters to their Planck 2015 values [24] to convert between the GW-measured luminosity distance and the cosmological redshift of the source. Changes to these parameters within the current range of uncertainties will not have any qualitative impact on our conclusions.

multaneously, because the detection efficiency of GW detectors depends on the component masses as well as the distance to the source, as discussed in §4.3. We parametrize the mass distribution as a power law with a variable upper mass cutoff, as in [105]. For the redshift distribution, we consider two parametrizations. The first parametrization is motivated by the low-redshift SFR, and assumes that the BBH merger rate follows the comoving volume to zeroth order in redshift. This model can fit astrophysically motivated redshift distributions, including metallicity-weighted SFRs convolved with various time-delay distributions. The second model allows for much more extreme deviations from a uniform in comoving volume merger rate, even at low redshifts, making it less suitable for distinguishing between different formation channels but more sensitive to the exotic scenarios discussed above, including modified gravity. These two models are described in §4.4.

In §4.5.1 we fit a joint mass-redshift distribution to the first six BBH detections announced by the LIGO-Virgo Collaboration (LVC): GW150914, LVT151012, GW151226, GW170104, GW170814, and GW170608 [21, 4, 5, 11, 13, 12]. For simplicity we treat LVT151012, which has an 87% probability of having astrophysical origin, as a full detection [3]. Although it has recently been suggested otherwise [51], we find that this set of detections is entirely consistent with a redshift distribution that is uniform in comoving volume. We show how future detections will improve the measurement of the BBH merger rate as a function of component masses and redshift in §4.5.2. We predict that with a few hundred BBH detections by LIGO-Virgo operating at design sensitivity, the measurement of the BBH redshift distribution will be precise enough to distinguish between different formation channels.

4.3 Detected redshift distribution

The distribution of redshifts among detected BBHs depends on the underlying mass distribution of the black holes. Assuming that the true BBH merger rate is constant in comoving volume, the redshifts of *detected* BBHs follow the cumulative probability distributions shown

in Figure 4.3, depending on their component masses. We assume detected BBHs are those that produce a signal-to-noise ratio (SNR) $\rho > 8$ in a single detector (see §4.5 for more details). Within the relevant mass range for stellar-mass BBHs, the GW signal from more massive BBH mergers is intrinsically stronger ("louder") and can be detected at greater distances. This means that the average redshift among detected heavy BBHs is higher than the average redshift among detected light BBHs. This can be seen in Figure 4.3, as the cumulative probability curves shift to the right with increasing BBH mass. Equivalently, the mass distribution of *detected* black holes is different from the true underlying mass distribution, with a preference for more massive black holes over less massive ones. Furthermore, as the sensitivity of the GW detector improves, the average redshift of the detected BBHs will increase. This is seen in the difference between the dashed curves, which assume a noise level appropriate to advanced LIGO's (aLIGO's) second observing run, and the solid curves, which assume the noise level for aLIGO at design sensitivity [respectively, the "Early High Sensitivity" scenario and "Design Sensitivity" noise curves from 7].

It is clear from Figure 4.3 that it is impossible to infer the underlying BBH redshift distribution independently of the BBH mass distribution. If the BBH merger rate density (rate per comoving volume) increases with increasing redshift, all of the cumulative probability curves in Figure 4.3 would shift to the right. However, increasing the relative number of massive BBH mergers in the population also increases the proportion of sources detected at high redshift. In other words, the measured redshift distribution alone cannot distinguish between a merger rate that increases with redshift and a population with a high fraction of massive BBHs [see also 34]. Likewise, the detected mass distribution is sensitive to the underlying redshift distribution; a greater fraction of detected BBHs will be high mass if the merger rate density increases with redshift, because only the high-mass BBHs are detectable at high redshift. In the following, we jointly examine the mass and redshift distribution of merging BBHs.

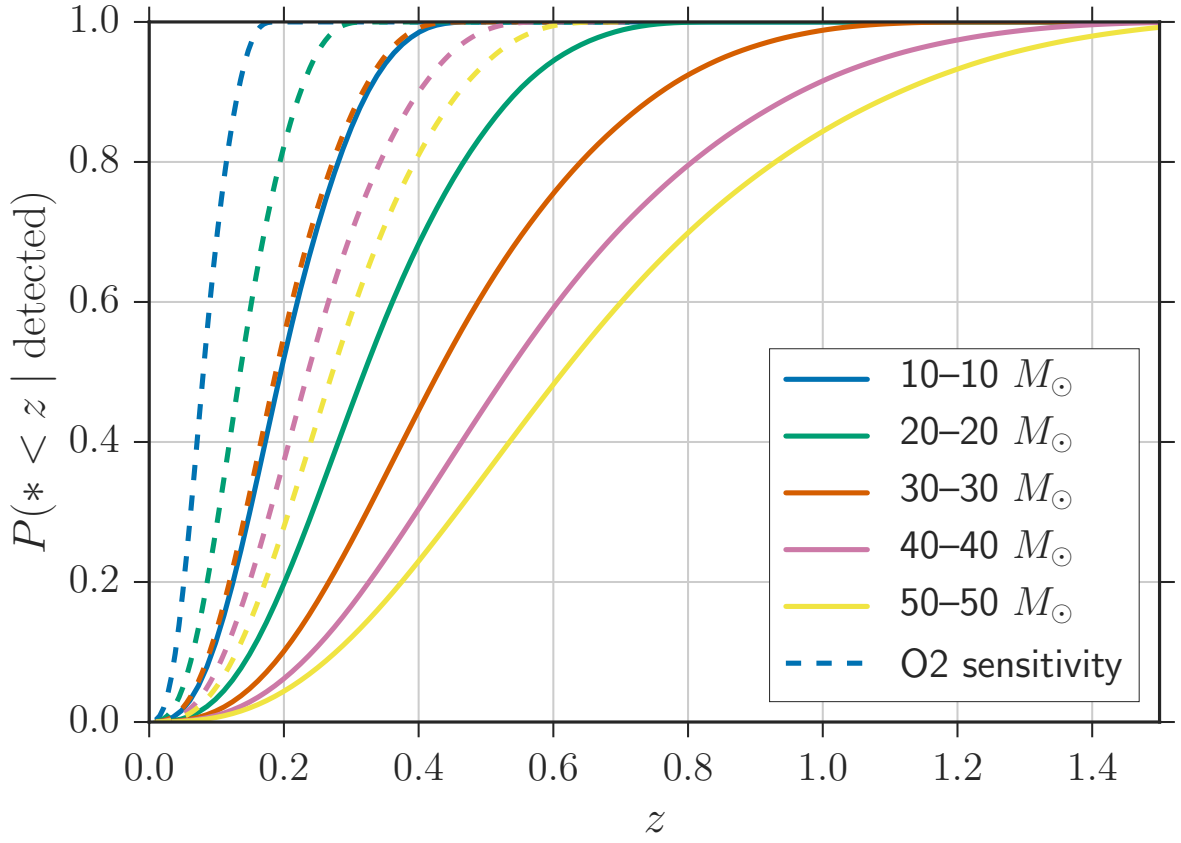

Figure 4.1: **Cumulative probability distribution of the redshifts of *detected* BBHs of given masses**, assuming that the underlying redshift distribution is uniform in comoving volume. The solid (dashed) lines show the expected distributions for aLIGO at design (O2) sensitivity. If the merger rate evolves positively (negatively) with redshift, these curves would shift to the right (left).

4.4 Joint Mass-redshift Model

We consider the differential mass-redshift distribution of BBHs:

$$\frac{dN}{dm_1 dm_2 dz} \equiv R\, p(m_1, m_2, z), \tag{4.1}$$

where R is the total number of BBHs across all masses and redshifts, so that:

$$\int \frac{dN}{dm_1 dm_2 dz} dm_1 dm_2 dz = R, \tag{4.2}$$

and $p(m_1, m_2, z)$ integrates to unity.

Given $\frac{dN}{dm_1 dm_2 dz}$, we can solve for the usual merger rate density, $\frac{dN}{dV_c dt_m}$, where t_m is the source-frame time. The merger rate density as a function of redshift is given by:

$$\frac{dN}{dV_c dt_m}(z) = \frac{dN}{dz}(z) \left[\frac{dV_c}{dz}(z)\right]^{-1} \frac{1+z}{T_{\rm obs}}, \tag{4.3}$$

where:

$$\frac{dN}{dz} = \int \frac{dN}{dm_1 dm_2 dz} dm_1 dm_2, \tag{4.4}$$

$T_{\rm obs}$ is the total observing time of the GW detector network as measured in the detector frame, and the $(1+z)$ factor converts detector-frame time to source-frame time.

As a first step, we assume that the underlying mass distribution does not vary across cosmic time, so that we can factor the joint mass-redshift distribution as:

$$p(m_1, m_2, z) = p(m_1, m_2) p(z). \tag{4.5}$$

This assumption may break down over a large range of redshifts, as many formation scenarios predict some dependence of the mass distribution on the merger redshift. However, aLIGO is only sensitive to redshifts $z \lesssim 1.5$ (see Figure 4.3), where Equation 4.5 is likely a good

approximation, particularly if the distribution of delay times is broad [see, for example, Figure 3 in 184, 65]. For the mass distribution, we use the two-parameter model from [105], which is an extension to the power-law model employed by the LVC to fit the BBH mass distribution to the first four detections [3, 11] incorporating the possibility of a mass gap above $\gtrsim 40\,M_\odot$ due to pair-instability supernovae [115, 127, 38]. We assume that the mass distribution takes the form:

$$p(m_1, m_2 \mid \alpha, M_{\rm max}) \propto \frac{m_1^{-\alpha}}{m_1 - 5\ M_\odot}\mathcal{H}(M_{\rm max} - m_1), \tag{4.6}$$

where $\mathcal{H}$ is the Heaviside step function. We fix the distribution of secondary masses, m_2, to be uniform between the minimum BH mass and m_1, and fix the minimum BH mass, $M_{\rm min} = 5\ M_\odot$.

4.4.1 Redshift Model A

For our first redshift model, we choose the following parametrization:

$$p(z \mid \lambda) \propto \frac{dV_c}{dz}\frac{1}{1+z}(1+z)^\lambda, \tag{4.7}$$

so that $\lambda = 0$ reduces to a merger rate density that is uniform in comoving volume and source-frame time. The extra factor of $(1+z)^{-1}$ converts from detector-frame to source-frame time. Note that for very small z, Equation 4.7 reduces to a constant in comoving volume and source-frame time merger rate regardless of the value of λ.

If the rate density follows the specific SFR, we would expect:

$$p(z) \propto \frac{dV_c}{dz}\frac{1}{1+z}\psi(z), \tag{4.8}$$

where $\psi(z)$ is the specific SFR [173]:

$$\psi(z) = 0.015 \frac{(1+z)^{2.7}}{1+[(1+z)/2.9]^{5.6}} \, M_\odot \, \mathrm{yr}^{-1} \, \mathrm{Mpc}^{-3}. \tag{4.9}$$

Other models for the SFR, such as [244] or [239], agree with the Madau-Dickinson SFR at the low redshifts relevant to aLIGO, $z < 1.5$. We note that Equation 4.7 with $\lambda = 2.7$ approximates Equation 4.8 for $z \ll 1$, whereas $\lambda = 2.4$ provides a very good approximation to Equation 4.8 for $0.1 \lesssim z \lesssim 1$. Alternatively, because BBH formation is more efficient at low metallicities, we might expect that the rate density follows the low-metallicity SFR, $\psi(z) f_Z(z)$, where $f_Z(z)$ is the fraction of star formation occurring at metallicity $\leq Z$ at redshift z [43, 182, 232, 2, 174]. For example, [166] give the fit:

$$f_Z(z) = \hat{\Gamma}\left(0.84, \frac{Z}{Z_\odot}^2 10^{0.3z}\right), \tag{4.10}$$

where $\hat{\Gamma}$ is the incomplete gamma function [see also 174]. As the average metallicity decreases with increasing redshift, the low-metallicity SFR rises more steeply with increasing redshift, and peaks at higher redshift. We find that a rate density that follows the low-metallicity ($Z \leq 0.3 Z_\odot$) SFR:

$$p(z) \propto \frac{dV_c}{dz} \frac{1}{1+z} \psi(z) f_{Z=0.3Z_\odot}(z), \tag{4.11}$$

leads to a redshift distribution, $p(z)$, that is well approximated by Equation 4.7 with $\lambda = 3.3$. It has also been proposed that the progenitors of BBHs are Population III stars formed at zero metallicity, in which case we might expect an even steeper increase of the merger rate with increasing redshift [36, 150].

More realistically, the rate density follows the SFR convolved with a time-delay distribution. Different formation channels predict different time-delay distributions. If typical time delays are very long (∼4–11 Gyr), as in the chemically homogeneous formation channel, the rate density will peak at very low redshifts ($z \sim 0.4$) well within the aLIGO horizon [174].

Under the parametrization of Equation 4.7, this corresponds to $\lambda < 0$; in the range $z \leq 1$, the best fits to such redshift distributions are given by $-6 \leq \lambda \leq -4$. In the classical field formation scenario, typical time delays are much shorter ($\sim$10–300 Myr) [88, 89]. In this field formation channel, the time delay is expected to follow a distribution:

$$\tau \propto \tau^{-1} \quad \tau_{\rm min} < \tau < \tau_{\rm max}, \tag{4.12}$$

where typically $\tau_{\rm min} \sim 50$ Myr and $\tau_{\rm max}$ is a Hubble time. In the redshift range of interest to aLIGO, this corresponds to a merger density that increases with increasing redshift ($\lambda > 0$), but is less steep than the SFR. For example, if the formation rate of BBHs follows the Madau-Dickinson SFR, and the time-delay between formation and merger follows Equation 4.12, the merger rate at $z \lesssim 1$ can be described by Equation 4.7 with $\lambda \sim 1.3$. A measurement of $\lambda > 1.3$ for the merger rate would indicate that the BBH formation rate density peaks at higher redshift than the SFR (possibly because of metallicity evolution), or that there is a stronger preference for very short time delays.

If the time-delay distribution is in fact restricted to very short time delays (for example, a flat distribution between $\tau_{\rm min} = 50$ Myr and $\tau_{\rm max} = 1$ Gyr), the BBH merger rate density will be nearly identical to their formation rate density at $z \lesssim 1$. This may be the case for mergers that take place inside globular clusters. On the other hand, binaries that form dynamically inside clusters but are ejected prior to merger tend to have much longer time delays, on the order of ~ 10 Gyr [211], corresponding to $\lambda \sim -10$.

4.4.2 Redshift Model B

For our second redshift model, we assume that the merger rate distribution is uniform in $V_c^{\gamma/3} = D_c^{\gamma}$ and source-frame time, so that $\gamma = 3$ implies that the distribution is uniform in

comoving volume. In other words, the redshift distribution takes the form:

$$p(z \mid \gamma) \propto \frac{1}{1+z} \frac{D_c^{\gamma-1}}{E(z)}, \tag{4.13}$$

where D_c is the comoving distance and $dD_c \propto \frac{dz}{E(z)}$ [133]. This redshift model is more flexible than Model A in the local universe, as it allows for large deviations from a constant in volume merger rate at low redshifts, whereas Model A always reduces to a constant in volume merger rate in the limit $z \to 0$. Model B can constrain scenarios that would cause extreme variations in the slope of the redshift distribution locally, such as a significant population of strongly lensed sources that appear closer than they are [51], leading us to infer $\gamma < 3$, or GW leakage causing sources to appear farther than they are, leading us to infer $\gamma > 3$ [80]. Previous studies have explored such scenarios through their effects on the SNR distribution [67, 118, 53] or GW standard siren measurements [203]. However, such effects on the redshift distribution (and likewise, the SNR distribution) will most likely be difficult to disentangle from the astrophysical processes that control the redshift and mass distributions. For example, in Figure 4.4.2, the dotted pink curve, corresponding to Redshift Model B, $\gamma = 4$, and the solid green curve, corresponding to Redshift Model A, $\lambda = 3$, are very similar.

Figure 4.4.2 shows the expected redshift probability density function (PDF) for sources detected by aLIGO at design sensitivity, assuming that the true mass-redshift distribution is described by the models discussed in this section. The solid, dashed, and dashed-dotted blue curves assume the same underlying redshift distribution (corresponding to a constant merger rate density), but different mass distributions. Meanwhile, the solid blue, orange, and green curves show how the detected redshift PDF varies with different underlying redshift distributions parametrized by λ, for a fixed mass distribution. The dotted pink curve assumes the same mass distribution, but takes the underlying redshift distribution to follow Model B with $\gamma = 4$. If the merger rate increases with redshift (for example, the solid green compared

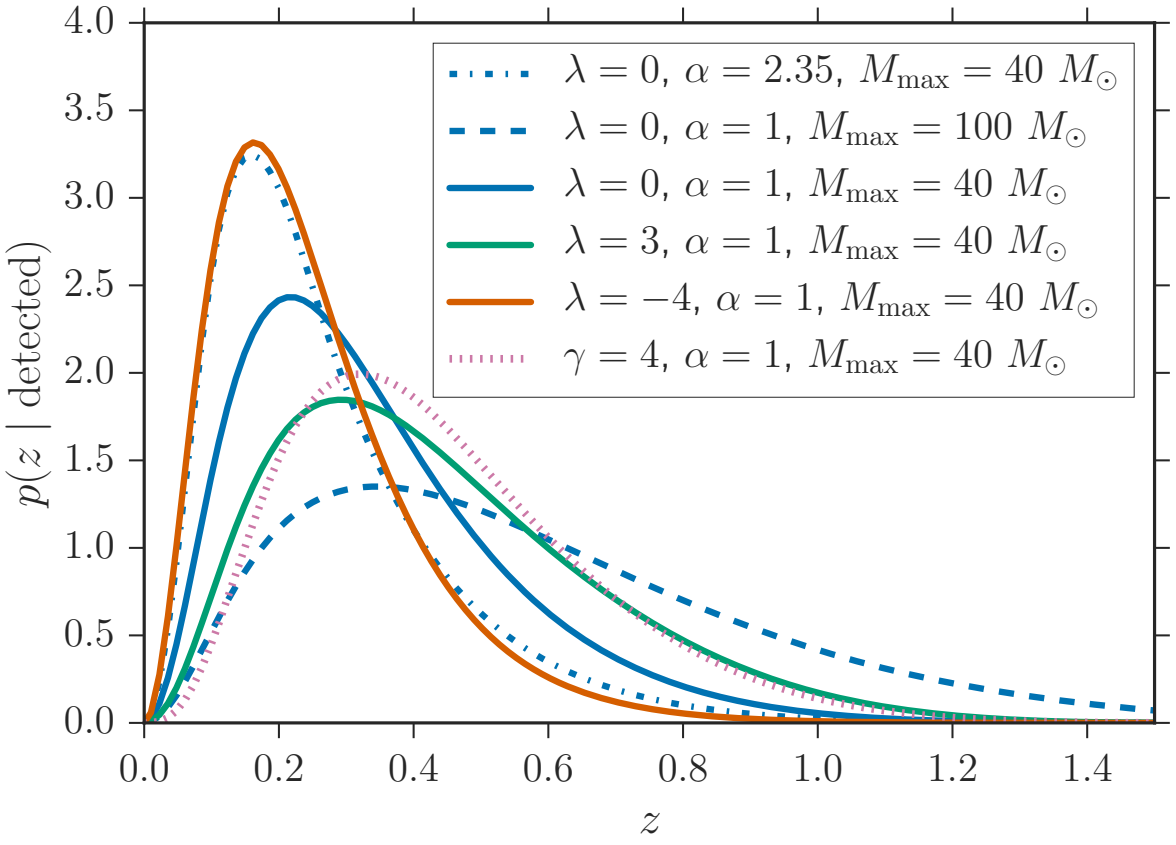

Figure 4.2: **Expected redshift distributions among the *detected* BBHs for LIGO-Virgo operating at design sensitivity**, for different choices of the underlying redshift distribution parametrized by λ (Model A) or γ (Model B). The detected redshift distributions depend on the underlying mass distribution, which we parametrize with a power-law slope, α, and an upper mass cutoff, $M_{\max}$.

to the solid blue curve), the detected distribution will skew to high redshifts. However, the effects of changing the mass distribution can be equally, if not more, significant (for example, the difference between the dashed, dashed-dotted and solid blue curves). As we shall see in the following section when we infer the parameters of the mass-redshift model, this leads to a degeneracy between the mass parameters and the redshift evolution parameter.

4.5 Fitting the Mass-redshift Distribution

In this section, we fit our parametrized model for the differential mass-redshift distribution, $\frac{dN}{dm_1 dm_2 dz}$, to real and simulated LIGO-Virgo detections. Our goal is to extract the four population parameters of the model from GW measurements of the masses and luminosity distances of detected sources. We assume a fixed ΛCDM cosmology determined by the 2015 Planck cosmological parameters [24], so that the measured luminosity distance is a direct measurement of the redshift. The shape of the mass-redshift distribution is governed by three parameters, $\boldsymbol{\theta}$. For Model A of the redshift evolution, $\boldsymbol{\theta} = \{\alpha, M_{\max}, \lambda\}$ and for Model B, $\boldsymbol{\theta} = \{\alpha, M_{\max}, \gamma\}$. The fourth parameter, R, corresponds to the total number of detected BBH systems and gives the overall normalization according to Equation 4.2, allowing us to solve for the physical merger rate of BBHs (Equation 4.3).

We model the rate density $\frac{dN}{dm_1 dm_2 dz}$ as a Poisson point process [169, 96, 255]. The likelihood for the GW data $\{d_i\}_{i=1}^{N_{\rm obs}}$ from $N_{\rm obs}$ observations, given population parameters

$\{\boldsymbol{\theta}, R\}$ is given by:

$$p(\{d_i\} \mid \boldsymbol{\theta}, R) = \left[\prod_{i=1}^{N_{\rm obs}} \int p(d_i \mid m_1, m_2, z) \frac{dN}{dm_1 dm_2 dz}\left(m_1, m_2, z \mid \boldsymbol{\theta}, R\right) dm_1 dm_2 dz\right] e^{-\beta(\boldsymbol{\theta}, R)} \tag{4.14}$$

$$= \left[\prod_{i=1}^{N_{\rm obs}} \left\langle \frac{1}{\pi(z^i, m_1^i, m_2^i)} \frac{dN}{dm_1 dm_2 dz}\left(m_1^i, m_2^i, z^i \mid \boldsymbol{\theta}, R\right)\right\rangle_{\{z^i, m_1^i, m_2^i\}}\right] e^{-\beta(\boldsymbol{\theta}, R)}, \tag{4.15}$$

where $\langle \cdots \rangle$ denotes an average over $\{z^i, m_1^i, m_2^i\}$ posterior samples from the ith event and $\pi(z^i, m_1^i, m_2^i)$ denotes the interim prior used in the analysis of individual events. The standard priors used in the LIGO analysis of individual events are uniform in component masses and "volumetric" in distance [245]:

$$\pi(z, m_1, m_2) \propto d_L(z)^2. \tag{4.16}$$

It is interesting to note that the redshift distribution described by Equation 4.16 matches Model A with $\lambda = 3$ (as opposed to $\lambda = 0$, which corresponds to a constant rate density), and so the simplifying assumption that the universe follows a Euclidean geometry implies a redshift distribution that mimics the SFR. Meanwhile, $\beta(\boldsymbol{\theta}, R)$ is given by:

$$\beta(\boldsymbol{\theta}, R) = \int \frac{dN}{dm_1 dm_2 dz}\left(m_1, m_2, z \mid \boldsymbol{\theta}, R\right) P_{\rm det}(m_1, m_2, z) dm_1 dm_2 dz \tag{4.17}$$

where $P_{\rm det}(z, m_1, m_2)$ is the fraction of binary sources at a given redshift and of given component masses that are detectable by the GW detector network. We assume that sources are isotropically distributed on the sky, and the binary inclination is uniformly distributed on the sphere (i.e., uniform in cos(inclination)). We also fix all BH spins to zero in our

analysis. Alternatively, we could allow the spins to vary and marginalize over the spin distribution when measuring the mass-redshift distribution [255]. However, incorporating the spin distribution will not affect our analysis significantly, considering that BBHs seem to have small aligned-spin components [99, 95].

Given the component masses and spins, sky position, inclination, and distance (or equivalently, redshift) of the BBH source relative to a GW detector, together with a power spectral density (PSD) that characterizes the noise of the detector, we can calculate the single-detector SNR [103, 91]. We consider a single-detector SNR threshold $\rho_{\rm th} = 8$ for detection (corresponding to a network SNR threshold of 12), and assume that for O1 and O2, the noise follows the PSD given by the aLIGO "Early High Sensitivity" scenario [7]. For this calculation we ignore the distinction between the true and measured SNR, which, with only six events, does not significantly impact our results. The probability of detection, $P_{\rm det}(z, m_1, m_2)$, is therefore the fraction of sources that produce a true SNR of $\rho = 8$ in a single detector.

We are interested in the posterior probability of the population parameters $\{\boldsymbol{\theta}, R\}$, which is related to the likelihood in Equation 4.14 by a prior:

$$p(\boldsymbol{\theta}, R \mid \{d_i\}) \propto p(\{d_i\} \mid \boldsymbol{\theta}, R) p(\boldsymbol{\theta}, R). \tag{4.18}$$

We choose broad, uninformative priors. We take a flat prior for the parameters that make up $\boldsymbol{\theta}$ (the power-law slope, maximum component mass, and redshift evolution parameter). Our default prior ranges are $\alpha \in [-4, 5]$, $M_{\rm max} \in [31\ M_\odot, 100\ M_\odot]$, $\lambda \in [-50, 30]$ and $\gamma \in [0, 8]$. For the rate parameter R, we take a flat-in-log prior over the range $R \in [10, 10^{12}]$. Recall that R is the total number of mergers between redshift $z = 0$ and the maximum redshift at which the detectors are sensitive—roughly $z = 0.6$ for O1 and O2 [2, 151]. The combined prior is then:

$$p(\boldsymbol{\theta}, R) \propto \frac{1}{R}. \tag{4.19}$$

With this prior choice, the posterior marginalized over R reduces to:

$$p(\boldsymbol{\theta} \mid \{d_i\}) \propto \prod_{i=1}^{N_{\rm obs}} \frac{\int p(d_i \mid m_1, m_2, z) p\left(m_1, m_2, z \mid \boldsymbol{\theta}\right) dm_1 dm_2 dz}{\xi\left(\boldsymbol{\theta}\right)} p(\boldsymbol{\theta}), \tag{4.20}$$

where:

$$\xi\left(\boldsymbol{\theta}\right) = \int p\left(m_1, m_2, z \mid \boldsymbol{\theta}\right) P_{\rm det}(m_1, m_2, z) dm_1 dm_2 dz \tag{4.21}$$

$$= \beta(\boldsymbol{\theta}, R)/R, \tag{4.22}$$

and $p\left(m_1, m_2, z \mid \boldsymbol{\theta}\right)$ is related to $\frac{dN}{dm_1 dm_2 dz}(m_1, m_2, z, \mid \boldsymbol{\theta}, R)$ by Equation 4.1. Equation 4.20 follows because Equation 4.18 can be written as:

$$p(\boldsymbol{\theta}, R \mid \{d_i\}) \propto p\left(\boldsymbol{\theta} \mid \{d_i\}\right) \left[\xi\left(\boldsymbol{\theta}\right)\right]^{N_{\rm obs}} R^{N_{\rm obs}-1} e^{-R\xi(\boldsymbol{\theta})}, \tag{4.23}$$

which when marginalized over R, yields $(N_{\rm obs}-1)!\, p\left(\boldsymbol{\theta} \mid \{d_i\}\right)$. Equation 4.20 is identical to the form of the posterior derived in [179] and used in previous population analyses of GW events [8, 105].

4.5.1 LIGO-Virgo detections

We fit our mass-redshift model to the first six announced BBH detections. We caution that at the time of writing, the analysis of LIGO's second observing run is still ongoing, and the sample of detections is not guaranteed to be complete. In order to avoid introducing any unmodeled selection biases, proper analysis should wait until the release of the final sample; nevertheless, our analysis illustrates the types of constraints we expect from six BBHs.

From each of these six events, we approximate the mass and redshift posterior PDFs using the published central values and 90% credible bounds. Specifically, we approximate the *detector*-frame chirp mass posterior PDFs by Gaussian distributions with a mean and

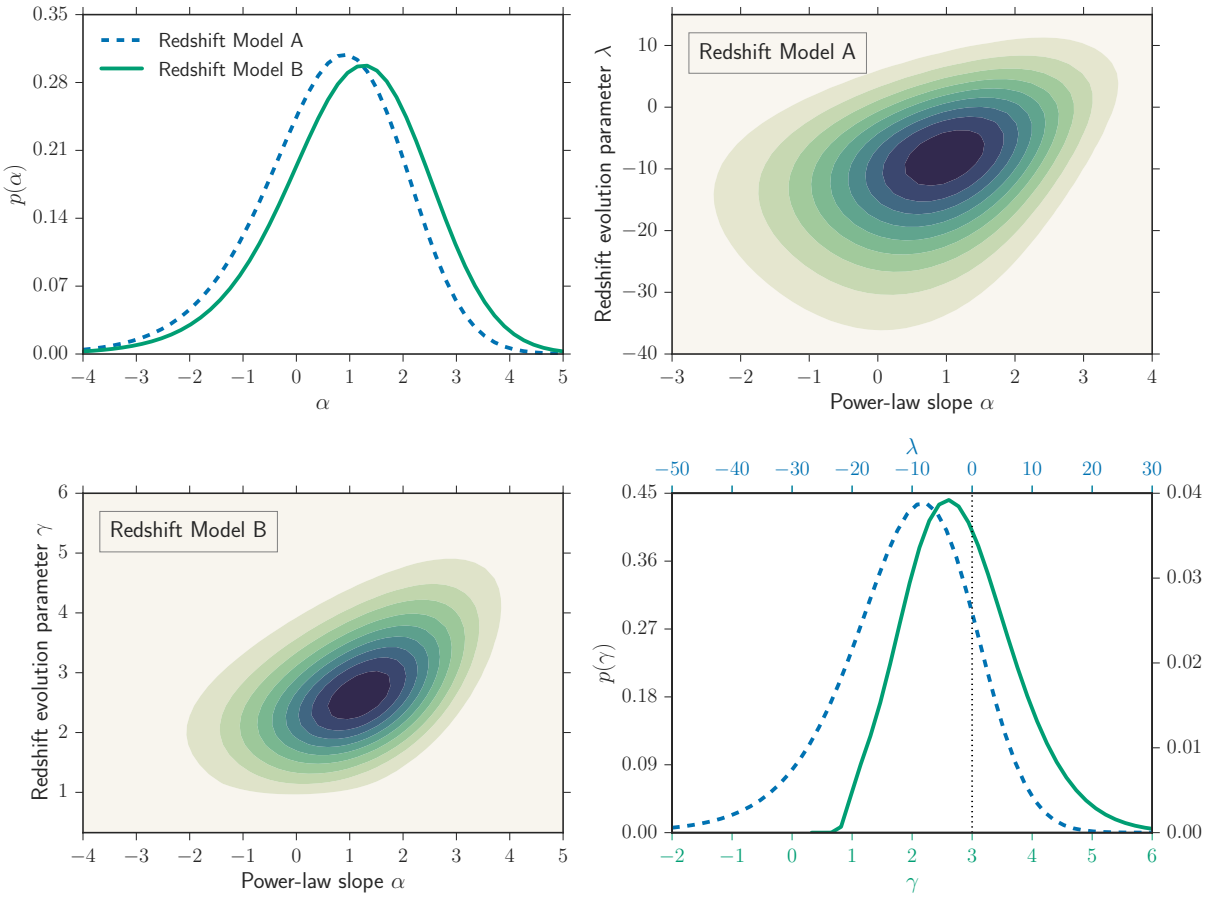

Figure 4.3: **Posterior PDF of the the power-law slope, α, and the redshift evolution parameter** from Model A (λ) and Model B (γ) from the first six announced BBH detections. The top right (bottom left) panel shows the two-dimensional posterior on α and λ (γ), calculated from the full posterior $p(\boldsymbol{\theta}, R \mid \mathbf{d})$ marginalized over $M_{\rm max}$ and R. The contours show increasing probability in 10% steps. The top left panel shows the posterior on α marginalized over all other parameters, for both Model A (dashed blue curve) and Model B (solid green curve) of the redshift evolution. The bottom right panel shows the posterior PDF for the redshift evolution parameters for the two models. The first six announced LIGO-Virgo detections are consistent with a uniform rate density ($\lambda = 0$ or $\gamma = 3$; dotted black line in bottom right panel) within the 68% credible interval, at the 56% (34%) credible level enclosing the maximum a posteriori value for Model A (Model B).

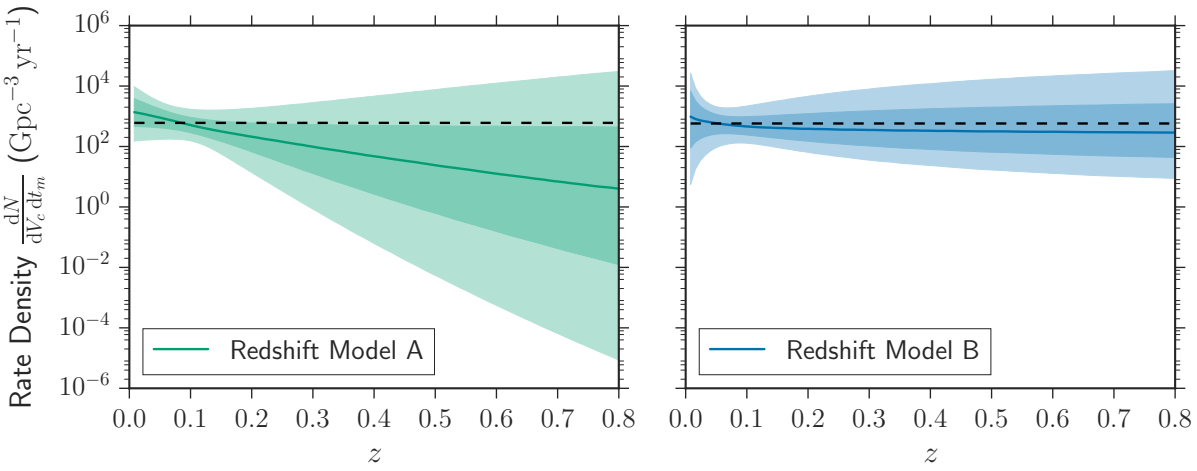

Figure 4.4: **Merger rate density as a function of redshift** for Model A (left) and Model B (right) of the redshift evolution, assuming that the six published LIGO-Virgo detections form a complete sample, and were detected during a 94-day observing period. The solid line shows the median rate density as a function of redshift, and the light and dark shaded regions show equal-tail 68% and 95% credible levels, respectively. Our inferred merger rate is consistent (at the 68% credible level) with being uniform in comoving volume and source frame time, t_m, which corresponds to a flat horizontal line on this plot (dashed black line). Our analysis shows a preference for a merger rate density that decreases with increasing redshift; however, this may be due to a false assumption that the six published BBHs form a complete sample from O1 and O2, as discussed in the text. Proper analysis, using the final sample and correct observing time, should wait for the analysis of O2 data to officially conclude.

standard deviation that match the published medians and 90% credible widths [3, 11, 13, 12]. The detector-frame masses differ from the source-frame masses by a factor of $(1+z)$ [161, 135]. Similarly, we use the published medians and 90% credible bounds on the mass ratio, $q = \frac{m_2}{m_1}$, to find the median and 90% credible bounds on the symmetric mass ratio:

$$\eta \equiv \frac{q}{(1+q)^2}. \tag{4.24}$$

We then approximate the symmetric mass-ratio posteriors by Gaussian distributions with means and standard deviations that match these medians and 90% credible intervals. We use these approximate posteriors on chirp mass and symmetric mass ratio to approximate the detector-frame component mass posterior distributions for each event. Lastly, we approximate the redshift posterior for each event by a Gaussian distribution matching the published median and 90% credible intervals. We therefore generate posterior samples for the detector-frame masses and redshifts following these approximate distributions. To get posterior samples for the *source*-frame masses, m_1 and m_2, we divide the posterior samples for the detector-frame masses by $(1+z)$, where the redshifts, z, are drawn from the redshift posterior distribution. This captures the correlations between redshift and source-frame masses in the posterior PDF for an individual event, $p(m_1, m_2, z \mid d_i)$.

Figure 4.5.1 shows the resulting posterior PDF on the power-law slope, α, and the redshift evolution parameter, λ or γ, marginalized over $M_{\rm max}$ and R:

$$p(\lambda, \alpha \mid \{d_i\}) = \int p(\lambda, \alpha, M_{\rm max}, R \mid \{d_i\}) dM_{\rm max} dR. \tag{4.25}$$

Under Model A, we infer $\lambda = -10^{+15}_{-21}$, $\alpha = 0.7 \pm 2.0$, and under Model B, we infer $\gamma = 2.7^{+1.8}_{-1.3}$, $\alpha = 1.1^{+2.1}_{-2.5}$. All credible intervals are quoted as the median and symmetric (equal-tailed) 90% range. Meanwhile, from the marginal posterior PDF on $M_{\rm max}$, we infer that the maximum component BH mass is 39^{+30}_{-6} $M_\odot$ (Model A) or 39^{+28}_{-6} $M_\odot$ (Model B);

the 95% upper limit of $\sim 69\ M_\odot$ is tighter than the 95% upper limit of $\sim 77\ M_\odot$ found in [105] with the additional two detections analyzed here.

There is a positive correlation between the mass power-law slope and the redshift evolution parameter in the two-dimensional posterior, because a mass distribution that favors low masses (large α) is compatible with the data only if the merger rate increases with redshift (large λ or γ) and vice versa; otherwise, more high-redshift and more low-mass objects would have been detected. The inferred rate parameter, R, is also positively correlated with these parameters, large α and/or λ (equivalently, γ) imply that there are many more high-redshift low-mass sources that contribute to the total number of mergers R, but not to the detected number, $N_{\rm obs}$ [see also 255].

It has recently been suggested that there are statistically too many nearby BBH detections (or equivalently, too many high-SNR detections) compared to the expected constant in comoving volume distribution [51]. Although our analysis shows a slight preference for a merger rate density that declines with increasing redshift, we find that for both models A and B of the redshift evolution, the current data is consistent with a uniform in comoving volume rate density ($\lambda = 0$ or $\gamma = 3$) within 1σ (68% credibility). Furthermore, it is possible that the set of published LIGO-Virgo detections is incomplete at this time. If loud events are published first, it is possible that an incomplete set would be biased toward low-redshift events, and our analysis would be artificially biased to small values of λ and γ. A more complete analysis can take place once the results from LIGO's second observing run are finalized.

Finally, we can calculate a posterior PDF on the merger rate density as a function of redshift according to Equation 4.3. As an illustration of the method, we assume that the total observing time from aLIGO's first and second observing runs is 94 days. This assumption is not based on the true observing time, which is not yet known as analysis on O2 is ongoing. Instead, it is chosen to match the most recently published merger rate estimate from LIGO

in [11], which is in the range [12, 213] $\mathrm{Gpc}^{-3}\ \mathrm{yr}^{-1}$ for a BBH population with a uniform in comoving volume merger rate and a mass distribution with power-law slope $1 \leq \alpha \leq 2.35$. With six published detections, an observing time of 94 days yields a mean merger rate of 100 $\mathrm{Gpc}^{-3}\ \mathrm{yr}^{-1}$ for the "power-law" ($\alpha = 2.35$) population considered in [11]. (Note that this "power-law" mass distribution from [11] fixes the maximum component BH mass to $M_{\mathrm{max}} = 95\ M_{\odot}$, the minimum component BH mass to $M_{\mathrm{min}} = 5\ M_{\odot}$, and the maximum total binary mass to $M_{\mathrm{tot,max}} = 100\ M_{\odot}$.) With this assumption of the observing time, Figure 4.5.1 shows the inferred rate density (marginalized over all population parameters) as a function of redshift for Model A (left panel) and Model B (right panel) of the redshift evolution. While previously published merger rate estimates are valid only for a fixed redshift distribution, this method allows us to infer the merger rate simultaneously with the mass and redshift distributions. Once again, we see that our results are consistent with a non-evolving merger rate, which would correspond to a flat horizontal line in Figure 4.5.1.

From Figure 4.5.1, we see that the merger rate is well-constrained at redshifts $0.05 \lesssim z \lesssim 0.15$. Meanwhile, there is too little volume at $z \lesssim 0.05$ to constrain the merger rate well, and the detectors are not sensitive enough at high redshifts. The uncertainties on the merger rate become especially large at high redshifts for Model A, because the parameter for this model, λ, is hard to constrain with low-redshift observations (where the merger rate always approaches a constant in comoving volume rate) but has a significant effect on the high-redshift rate. Meanwhile, varying the parameter, γ, in Model B causes a large variation in the low-redshift rate. This means that γ is easier to measure with low-redshift observations than λ, and so assuming Model B, the merger rate is relatively well constrained at high redshifts where the sensitivity of the GW detectors approaches zero.

4.5.2 Future detections

In this section, we simulate detections from a mock population of BBHs and apply our method to infer the underlying mass and redshift distribution parameters. Our goal is to estimate how many LIGO-Virgo detections will be required to correctly infer a deviation from a uniform in comoving volume merger rate, or alternatively, how many detections will be required to confidently rule out strong deviations from a uniform in comoving volume merger rate. We consider two simulated populations. Both populations follow the same distribution for the BBH masses (Equation 4.6), with a minimum component mass of 5 $M_\odot$, a maximum component mass of 40 $M_\odot$, and a power-law slope $\alpha = 1$. We assume that the mass distribution for both populations is independent of the redshift distribution (Equation 4.5). The redshift distribution of the first population follows Model A (Equation 4.7) with $\lambda = 3$, as might be expected if the merger rate followed the low-metallicity SFR convolved with a time-delay distribution that favored short time delays. Meanwhile, the second population has a uniform in comoving volume merger rate, corresponding to $\lambda = 0$ in Model A or $\gamma = 3$ in Model B.

For each mock population we generate BBHs with component masses and redshifts following the assigned underlying distribution. Only a subset of BBHs are detected, and their measured masses and redshifts take the form of (marginalized) posterior PDFs. In order to generate realistic mass and redshift measurements, we construct a synthetic detection model. The synthetic detection model enables us to self-consistently and realistically capture the correlations between the measured SNR, which determines the detectability of an event and its measured redshift. We verify that the model closely approximates the detectability of a BBH with given masses and redshift by aLIGO at design sensitivity.

Synthetic BBHs are generated as follows. Each BBH system is characterized by four parameters: the source-frame chirp mass $\mathcal{M}$, the symmetric mass ratio η, the luminosity distance d_L, and an angular factor Θ. Each system also has an associated *true* SNR, ρ,

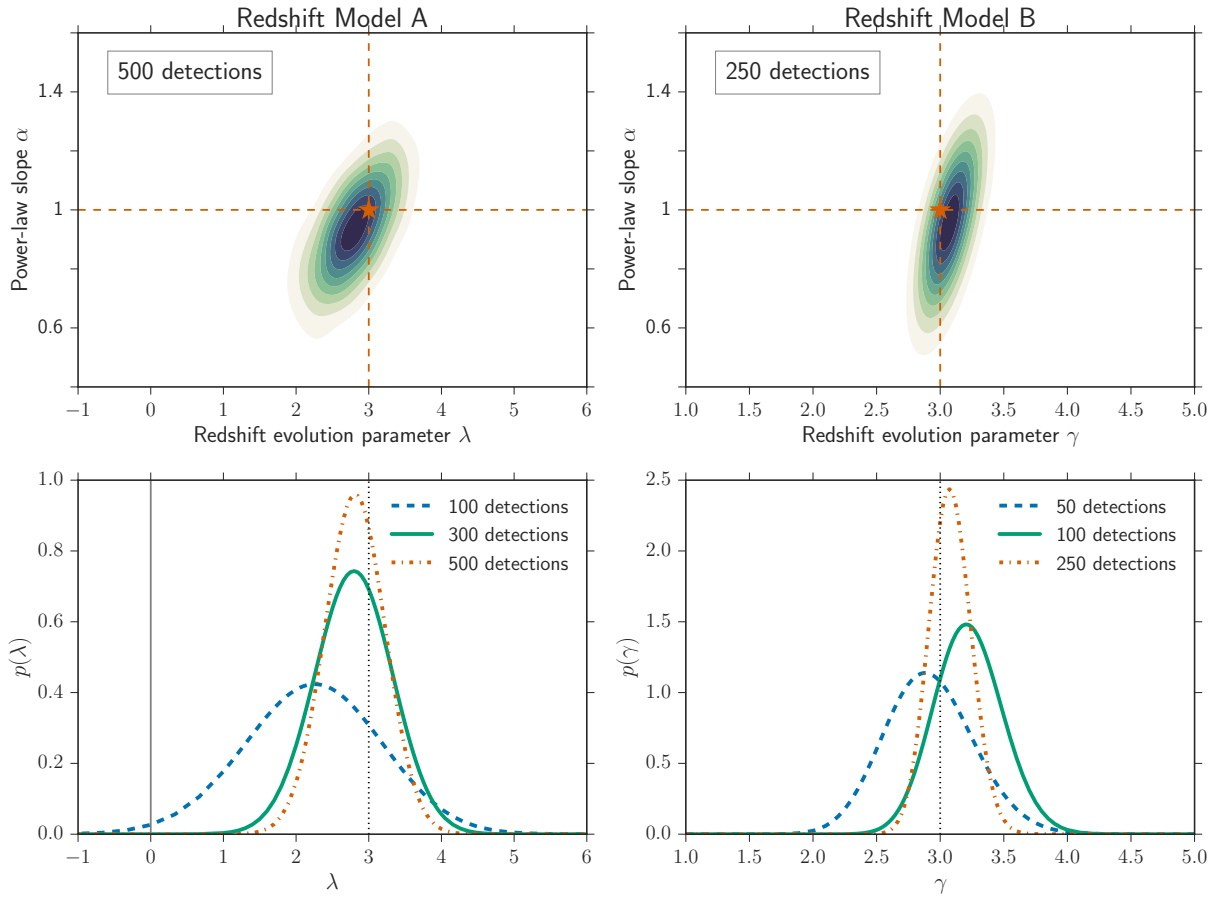

Figure 4.5: **Projected constraints on the mass power-law slope and the redshift evolution parameter for a set of simulated BBH detections from two populations.** The top panels show the joint posterior PDF on the power-law slope and redshift evolution parameter, marginalized over the rate and maximum mass parameters. The bottom panels show the marginalized posterior on the redshift evolution parameter. Both populations follow the same mass distribution described by $\alpha = 1$, $M_{\rm max} = 40\ M_\odot$, but differ in their redshift distribution. *Left panels:* This population is described by a redshift distribution that roughly follows the SFR, or $\lambda = 3$ in Model A. After 100 detections by LIGO-Virgo at design sensitivity (solid green line, bottom left panel), the constraints on λ are tight enough to exclude a uniform in comoving volume merger rate, $\lambda = 0$ (black solid line), at 99% credibility. *Right panels:* This population has a uniform in comoving volume merger rate ($\gamma = 3$ in Model B). We fit detections from this population with Model B of the redshift evolution, and the parameter γ is sufficiently well-constrained after 100 detections to constrain γ to a 90% credible interval of $\lesssim 1$.

which depends on these four parameters. Note that $\mathcal{M}$ and η allow us to directly infer m_1 and m_2, and d_L allows us to infer z. Here, Θ plays the combined role of the sky location, inclination, and polarization on the measured GW amplitude. We tune the width of the Θ distribution to control the uncertainty of the *measured* signal strength, which in turn controls the uncertainty on the measured luminosity distance. This allows us to capture the correlations between the measured signal strength and the measured redshift, which is necessary in order to model the selection effects consistently between the detection model and the calculation of $P_{\rm det}$ (Equation 4.17).

We set the "typical" SNR, ρ_0, of a BBH system with parameters $\mathcal{M}$ and d_L to:

$$\rho_0 \equiv 8 \left(\frac{\mathcal{M}\,(1+z)}{\mathcal{M}_8} \right)^{5/6} \frac{d_{L,8}}{d_L}, \tag{4.26}$$

where we fix $\mathcal{M}_8 = 10\ M_\odot$ and $d_{L,8} = 1$ Gpc. This scaling approximates the amplitude of an inspiral GW signal to first order, and we chose $\mathcal{M}_8$ and $d_{L,8}$ to roughly match the typical distances of detected sources by aLIGO at design sensitivity [71].[2] The true SNR, ρ, in our model depends on the angular factor Θ, and is given by:

$$\rho = \rho_0 \Theta, \tag{4.27}$$

similar to relationship between the true SNR and the "optimal SNR" via the projection factor Θ [103] or w [91, 71], although Θ in our case is simply a random variable with a log-normal distribution. As the uncertainty on Θ controls the uncertainty on d_L, we pick the variability of Θ in order to get realistic measurement uncertainties on d_L. We find that realistic measurement uncertainties on d_L are achieved when Θ has a typical width of 15%,

2. It should be noted that this scaling breaks down for high-mass sources, where a significant fraction of the SNR comes from the merger and ringdown components of the signal, rather than the inspiral. However, the heaviest BBHs in our simulated population are 40–40 $M_\odot$ in the source frame, and so our synthetic model provides a good approximation to their detectability.

and so we pick:

$$\log \Theta \sim N\left(0, \frac{0.3}{1+\frac{\rho_0}{8}}\right). \tag{4.28}$$

From the true parameters $\mathcal{M}$, η, d_L and Θ, we assume that the measurement process measures three parameters: the observed SNR, $\rho_{\rm obs}$, the observed chirp mass, $\mathcal{M}_{\rm obs}$, and the observed symmetric mass ratio, $\eta_{\rm obs}$. These are given by:

$$\rho_{\rm obs} = \rho + N(0,1), \tag{4.29}$$

$$\log \mathcal{M}_{\rm obs} = \log\left(\mathcal{M}(1+z)\right) + N\left(0, 8\frac{\sigma_{\mathcal{M}}}{\rho_{\rm obs}}\right), \tag{4.30}$$

$$\eta_{\rm obs} = \eta + N\left(0, 8\frac{\sigma_{\eta}}{\rho_{\rm obs}}\right), \tag{4.31}$$

where we assume that the observed SNR is normally distributed about the true SNR, ρ, with a standard deviation of 1 (due to different realizations of Gaussian noise), and the uncertainties on the mass parameters scale inversely with the observed SNR [246]. We fix a threshold of $\rho_{\rm obs} \geq 8$ for detection. To match the expected measurement uncertainties, we fix $\sigma_{\mathcal{M}} = 0.04$ and $\sigma_{\eta} = 0.03$, so that the relative 90% credible interval uncertainty for the recovered detector-frame primary (secondary) mass is typically 40% (50%) [121, 251]. Meanwhile, the luminosity distance is measured from $\rho_{\rm obs}$ via Equations 4.26–4.28. The typical relative 90% confidence interval uncertainty for the recovered luminosity distance is $\sim 50\%$, which is also a realistic expectation [251]. As discussed earlier in this section, this process of recovering the measured luminosity distance from the measured signal strength is necessary in order to incorporate selection effects consistently, and ensure that we generate single-event posteriors, $p(m_1, m_2, z)$ that are compatible with the assumed detection probability, $P_{\rm det}(m_1, m_2, z)$.

Under this synthetic (yet realistic) detection model, we generate 500 *detected* BBH systems for each of the two populations. The projected constraints on the power-law slope and redshift evolution parameter are shown in Figure 4.5.2. Note that the maximum mass

parameter will be already tightly measured with a few tens of detections [105]. We find that after 100 detections by LIGO-Virgo (which may happen as early as the next observing run, starting in late 2018) it may be possible to detect deviations from a uniform in comoving volume merger rate if the true redshift distribution evolves as steeply as the low-metallicity SFR ($\lambda \sim 3$). (This is expected from formation channels where the typical time delay between formation and merger is short.) Additionally, we expect to distinguish between a merger rate density that increases with increasing redshift and a merger rate density that decreases with redshift (as expected from formation channels with very long time delays). If the true deviation from a uniform merger rate density is small ($0 < \lambda < 1$, where $\lambda = 0$ implies a uniform merger rate density), it may take ~ 500 detections to confidently exclude $\lambda = 0$. This will require a few years of aLIGO operating at design sensitivity [starting in 2020+ 7]. Meanwhile, extreme deviations from a constant merger rate density ($\gamma \neq 3$ in Redshift Model B) can be ruled out in 100 detections, as γ will be constrained to a 90% credible interval width of $\lesssim 1$.

In summary, we expect that with N detections by LIGO-Virgo operating at design sensitivity, we will be able to constrain λ from Model A to a 90% credible interval of width $\sim 31/\sqrt{N}$ and γ from Model B to a 90% credible interval of width $\sim 8.5/\sqrt{N}$, although the exact rate of convergence depends on the true values of these parameters. If the local BBH merger rate is 100 Gpc^{-}3 yr^{-1}, the mass distribution follows a power law with $\alpha = 1$ and $M_{\rm max} = 40\ M_{\odot}$, and the redshift distribution is constant in comoving volume in source-frame time, we expect ~ 300 detections in one year of LIGO/Virgo operating at design sensitivity (assuming that each detector has a duty cycle of 80%, and that a confident detection requires at least two detectors in observing mode). Fixing this mass distribution, but assuming instead that the redshift distribution follows the Madau-Dickinson SFR, the expected number of detections increases to ~ 680. On the other hand, if the overall merger rate is the same, but the BBH mass distribution follows a steeper power law with $\alpha = 2.35$

(fixing $M_{\rm max} = 40\ M_\odot$), we expect ~ 140 detections in a year of LIGO/Virgo operating at design sensitivity if the merger rate density is independent of redshift, and ~ 280 detections if the merger rate density follows the SFR. (For O3 sensitivity, the expected number of detections is smaller by a factor of 4–5.) With 100–700 detections per year, we expect to detect deviations from a constant merger rate density, or severely constrain such deviations, within the first 1–3 years of LIGO/Virgo operating at design sensitivity (starting ~ 2020).

4.6 Discussion

We have explored the ability of the LIGO-Virgo network to measure the redshift evolution of the BBH population. We have applied a simple four-parameter model to constrain the BBH merger rate density, $\frac{dN}{dm_1 dm_2 dz}$, as a function of component masses and redshift. Our model allows us to simultaneously constrain the slope and maximum mass of the distribution of primary BH masses, the slope of the redshift distribution, and the merger rate. We note that our method can also be applied to the binary neutron star (BNS) population, although such sources will only be detectable up to redshifts $z < 0.1$ with the current generation of GW detectors. However, the mass distribution of such sources may already be well constrained from the galactic population [201]. If we adopt this mass distribution as a prior, the analysis on BNS would simplify to a one-parameter redshift evolution model.

Recall that a measurement of the merger redshift distribution constrains a combination of the formation rate as a function of redshift and the time-delay distribution. If we can constrain the redshift evolution parameter to $\lambda \gtrsim 2.4$, we may infer that the BBH formation rate density peaks at higher redshift than the SFR, regardless of the time-delay distribution. If we assume that the time-delay distribution follows Equation 4.12 with $\tau_{\rm min} = 50$ Myr and $\tau_{\rm max} = 14$ Gyr, a measurement of $\lambda \gtrsim 1.3$ implies that the BBH formation rate density peaks at higher redshift than the SFR. Alternatively, if we assume that the formation rate density follows the low-metallicity SFR in Equation 4.11, measuring $\lambda \gtrsim 1.9$ ($\lambda \lesssim 1.9$) would

allow us to infer that the time-delay distribution is more skewed towards short (long) time delays than Equation 4.12.

Because our focus is on extracting the redshift evolution of the merger rate, we have simplified our treatment of the mass distribution. For example, our parametrization assumes that the BBH mass distribution does not evolve with redshift, and does not allow the distribution of mass-ratios or the minimum BH mass to vary. We have verified that adding two free parameters, β and $M_{\rm min}$, to describe the mass-ratio distribution and the minimum mass according to:

$$p(m_1, m_2 \mid \alpha, \beta, M_{\rm max}, M_{\rm min}) \propto \frac{m_1^{\alpha} m_2^{\beta}}{m_1 - M_{\rm min}} \mathcal{H}\left(M_{\rm max} - m_1\right), \tag{4.32}$$

does not significantly affect our results for the six BBH detections. Taking uniform priors on these additional parameters, $-4 < \beta < 4$, $3 < M_{\rm min} < 9\ M_{\odot}$ [3], causes the posteriors on the remaining four parameters to widen only slightly. Furthermore, the constraints on the additional two parameters, β and $M_{\rm min}$, are not informative with only six detections, and are consistent with our default values, $\beta = 0$, $M_{\rm min} = 5\ M_{\odot}$, although we find a slight preference for equal mass ratios ($\beta > 0$ at $\sim 70\%$ credibility; consistent with the results of [215]) and a larger minimum mass ($M_{\rm min} > 3.9\ M_{\odot}$ at 95% credibility). These results hold for both Models A and B of the redshift evolution. It will likely take $\mathcal{O}(100)$ detections to measure $M_{\rm min}$ sufficiently well and resolve the putative gap between the neutron star and BH mass spectrum [168, 175, 159].

With sufficient detections, our four-parameter model will likely break down, and we should include more degrees of freedom in the mass-redshift (and possibly also spin) distribution to avoid introducing systematic biases in the inferred parameters [240, 255]. For example, if the mass distribution varies with redshift, possibly favoring larger masses at high

3. The minimum BH mass is constrained to be no larger than $\sim 9\ M_{\odot}$, the 95% upper bound on the secondary mass of GW170608 [12].

redshifts due to the lower average metallicity, our simple model will misinterpret this as an evolution in the merger rate. Therefore, a more complicated model should allow for correlations between the mass and redshift distribution, either through the addition of one to two parameters (e.g. a copula model), or a many-parameter model that fits the mass distribution separately in different redshift bins. It may also be possible to introduce a multi-component mixture model to determine whether there are multiple populations of BBHs following different mass-redshift distributions. Furthermore, a more sophisticated model would include at least two additional parameters to fit the peak, $z_{\rm peak}$, and the high-redshift slope of the merger rate density. For example, we can consider the following parametrization of the merger rate density inspired by the Madau-Dickinson SFR:

$$p(z \mid a, b, z_{\rm peak}) \propto \frac{1}{1+z}\frac{dV_c}{dz}\frac{(1+z)^a}{1+\left(\frac{a}{b-a}\right)\left[\frac{1+z}{1+z_{\rm peak}}\right]^b}. \tag{4.33}$$

This parametrization provides an excellent fit to all of the astrophysical redshift distributions discussed in Section 4.4 up to redshifts $z \sim 4$, whereas our one-parameter model of Equation 4.7 starts to break down at $z \sim 1$. However, it is unlikely that we will have tight constraints on $z_{\rm peak}$ and b with second-generation GW detectors, because $z_{\rm peak}$ will likely lie beyond the sensitivity of these detectors, unless the time delays between formation and merger are typically extremely long (greater than a few Gyr). For example, if the BBH formation rate follows the Madau-Dickinson SFR and the time-delay distribution in Equation 4.12, the peak of the merger rate density would be at $z_{\rm peak} = 1.4$, where we expect to have very few detections (see Figures 4.3 and 4.4.2). If the BBH formation rate peaks later than the Madau-Dickinson SFR, following the low-metallicity SFR in Equation 4.11 for example, the peak of the merger rate density would be even farther out of reach, at $z_{\rm peak} = 2.1$, assuming the same time-delay distribution. However, by the time we have $\mathcal{O}(1000)$ detections, it may be possible to get some measurement of $z_{\rm peak}$, which would allow us to infer the peak of

the formation rate density for an assumed time-delay distribution (this peak would be at $z = 1.8$ if BBH formation followed the Madau-Dickinson SFR, and $z = 2.7$ if the formation followed the low-metallicity SFR of Equation 4.11). We look forward to a time where our single redshift-evolution parameter is sufficiently well measured that we must include these additional parameters in our model; we anticipate that this will take over 500 detections. We note that with third-generation GW detectors, it will be possible to accurately infer the entire formation rate history of BBHs together with the time-delay distribution from the observed redshift evolution of the merger rate [250].

In addition to the limitations of our parametrized model, another much less significant source of systematic uncertainty in our analysis comes from GW measurements of the luminosity distance (and therefore, the redshift) to a source. Extracting the luminosity distance from a GW signal depends on measuring its amplitude, which is affected by detector calibration uncertainties. The calibration uncertainty is only a few percent [144], which is negligible compared to the expected uncertainty on the redshift evolution parameter. Another subdominant source of uncertainty comes from the effect of weak lensing on the GW amplitude, which contributes at the sub-percent level and is therefore negligible for our analysis [137, 136].

4.7 Conclusion

By fitting a four-parameter mass-redshift distribution to the first six announced BBH mergers, we have placed the first constraints on the redshift evolution of the BBH merger rate. We show that because of strong correlations between the masses and redshifts of detected BBHs, the mass and redshift distribution must be fit simultaneously. We consider two parametrizations of the redshift evolution: Model A fixes the slope of the redshift distribution to match the differential comoving volume locally (as $z \to 0$), as is expected from most astrophysical formation channels, while Model B allows for large deviations, even at low redshift. Our

constraints from six events are too weak to distinguish between any astrophysical formation scenarios: for example, we measure $-31 < \lambda < 5$ at 90% credibility for Model A, whereas typical formation channels predict between $-4 \lesssim \lambda \lesssim 3$). However, we can already constrain extreme deviations from a uniform in comoving volume merger rate, finding $\gamma = 2.7^{+1.8}_{-1.3}$ for Model B ($\gamma = 3$ corresponds to a uniform merger rate). Furthermore, while previous analyses have calculated the BBH merger rate under the assumption of a uniform in comoving volume redshift distribution, we demonstrate how to infer the merger rate density as a function of redshift.

We project that with 100–500 detections by LIGO-Virgo, the inferred redshift evolution of the BBH merger rate will allow us to distinguish between proposed formation channels (for example, those that favor long versus short time delays between progenitor formation and BBH merger). Meanwhile, the overall merger rate and mass distribution will also provide important clues regarding the formation channel [91, 237, 256, 35]. In 5–10 years, the constraints on the redshift evolution parameter alone will allow us to infer the peak of the formation rate of BBH progenitors and/or the typical time delay between formation and merger.

CHAPTER 5

PICKY PARTNERS: THE PAIRING OF COMPONENT MASSES IN BINARY BLACK HOLE MERGERS

As we saw in Chapter 3, the masses of individual black holes contain astrophysical clues to their formation histories. The absence of component black holes with masses greater than 40 $M_\odot$, for example, is significant for learning about supernovae physics and for distinguishing black holes with stellar progenitors from those formed from previous mergers. In this chapter, we focus on the two masses that make up a merging binary black hole system, and examine the relationship between them.

Different formation scenarios for binary black hole formation generally predict some relationship between the component masses. For example, the chemically homogeneous formation channel can only produce symmetric systems in which the two components have the same mass to within $\sim 90\%$, and certain dynamical formation channels predict that the probability of two black holes in a globular cluster finding each other and merging scales with the total mass of the system. Motivated by theoretical predictions, we consider a simple phenomenological model of the two-dimensional mass distribution, parameterized as a product of the single black hole mass function and the *pairing function*. We allow the pairing function to depend on the mass ratio between the two components and on the total mass of the binary system. If the pairing function is independent of the masses, we say that the pairing between the component black holes is random.

Fitting the pairing function to the ten binary black holes from LIGO and Virgo's first two observing runs, we find that the pairing between component masses is unlikely to be random. Rather, we infer a statistical preference for mergers between equal-mass components, suggesting that black holes are picky about their partners. While the mass ratio likely affects the pairing probability, we do not find evidence that the pairing function depends on total mass. Comparing the inferred pairing function to theoretical predictions, we can start

to probe the physics of binary interactions.

This work was carried out in collaboration with Daniel Holz, and published in the Astrophysical Journal Letters [108]. Shortly after this work was completed, LIGO and Virgo detected the first confidently asymmetric binary black hole merger: GW190412 [22]. This binary consists of a $\sim 30\ M_\odot$ and $\sim 9_\odot$. Using the analysis of this chapter, we are able to label GW190412 as an "exceptional" event. Because the first ten detections suggest a strong underlying preference for symmetric (equal-mass) mergers, the detection of a confidently asymmetric merger is surprising, and may challenge binary formation scenarios.

MF was supported by the NSF Graduate Research Fellowship Program under grant DGE-1746045. MF and DEH were supported by NSF grant PHY-1708081. They were also supported by the Kavli Institute for Cosmological Physics at the University of Chicago through NSF grant PHY-1125897 and an endowment from the Kavli Foundation. DEH also gratefully acknowledges support from the Marion and Stuart Rice Award.

5.1 Abstract

We examine the relationship between individual black hole (BH) masses in merging binary black hole (BBH) systems. Analyzing the ten BBH detections from LIGO/Virgo's first two observing runs, we find that the masses of the component BHs comprising each binary are unlikely to be randomly drawn from the same underlying distribution. Instead, the two BHs of a given binary prefer to be of comparable mass. We show that it is ~ 5 times more likely that the component BHs in a given binary are always equal (to within 5%) than that they are randomly paired. If we assume that the probability of a merger between two BHs scales with the mass ratio q as q^β, so that $\beta = 0$ corresponds to random pairings, we find $\beta > 0$ is favored at credibility 0.987. By modeling the mass distribution, we find that the median mass ratio is $q_{50\%} = 0.91^{+0.05}_{-0.17}$ at 90% credibility. While the pairing between BHs depends on their mass ratio, we find no evidence that it depends on the total mass of the system:

it is ~ 6 times more likely that the pairing depends purely on the mass ratio than on the total mass. We predict that 99% of BBHs detected by LIGO/Virgo will have mass ratios $q > 0.5$. We conclude that merging black holes do not form random pairings; instead they are selective about their partners, preferring to mate with black holes of a similar mass. The details of these selective pairings provide insight into the underlying formation channels of merging binaries.

5.2 Introduction

Following the first two observing runs (O1 and O2) of advanced LIGO [1] and Virgo [23], the LIGO/Virgo Collaboration (LVC) reported ten detections of merging binary black holes (BBH) [17], with tens more detections expected from the third observing run (O3), and hundreds of expected detections per year once the gravitational-wave (GW) detector network reaches design sensitivity [7]. In addition to the LVC-published detections of [17], new BBH detections in the O1 and O2 data have been reported by [248], [247], and [196]. The formation and history of these BBHs remains a fundamental question in GW astrophysics. The proposed formation channels include isolated [91, 39, 253, 94, 45, 238, 165, 233], dynamical [183, 139, 212, 32, 65, 210, 218, 257, 83], and primordial [48, 117], with many variants within each model. Different formation channels are expected to leave an imprint on the properties of the BBH population [35, 241, 30], including the mass distribution [237, 256], spin distribution [214, 100, 252, 95], and redshift evolution [109, 250, 213]. It is therefore possible to learn about the astrophysics of BBH formation by fitting for these population distributions using GW data. In [19], the LVC carried out such an analysis on the first ten BBH detections, fitting the mass, spin and redshift distributions with simple parameterized models. For example, the mass distribution was fit to a model in which the primary mass (the more massive component of a binary) follows a power-law between some minimum and maximum mass, while the secondary mass is distributed with a power-law between the

minimum mass and its primary mass partner. [19] additionally considered a slightly more complex model, which replaces the minimum mass cutoff with tapering at the low-mass end and allows for an additional Gaussian component at the high mass end of the primary-mass power-law. In this work, we restrict the population analysis to the ten [17] BBH detections, as the detection efficiency has been previously studied for this sample and is well-understood. The detection efficiency between the LVC detections and the [247] detections is significantly different; see e.g. Figure 5 of [247]. Using the wrong detection efficiency leads to selection biases in population inference. In future work we will extend our analysis to include overlapping samples with differing selection effects.

In this work we extend the analysis of [19] by focusing on a particular aspect of the BBH mass distribution: the pairing between the two component BHs in the binary. We ask whether the universe makes merging binary black hole systems by randomly pairing up black holes, or whether the mass of each black hole in a pair influences the mass of its companion. This differs from the analysis of [19], in which the parameterization for the mass distribution does not separate the underlying BH mass distribution and the pairing function. Under the models considered by [19], it is not possible to fit for an underlying mass distribution that is common to both component BHs or quantify the deviation from the random-pairing scenario, as we do in this work.

We expect that the pairing function carries an imprint of the physics by which component BHs find their partners. Various formation models predict that the mass ratio and/or total mass of the two components may determine their probability of merging. Despite the different physical processes involved, many formation channels predict a preference for similar component masses [177]. Binaries formed via homogeneous chemical evolution are expected to strongly prefer equal mass components due to the progenitor stars exchanging mass during an early overcontact phase [174, 186]. The traditional isolated evolution channel is also expected to favor comparable mass components, because the common envelope phase is unsuc-

cessful at producing close binaries for extreme mass ratio systems [91]. However, the common envelope phase remains poorly understood, and this channel can produce mergers between fairly unequal component masses, especially at lower metallicities [90, 238, 152, 122, 234]. Some studies have suggested that dynamical evolution also tends to produce more mergers with equal mass components due to the fact that comparable mass binaries have a higher binding energy and form tighter binaries [211, 27]. However, other dynamical channels may mildly prefer unequal mass components [190]. Alternatively, it has been suggested that in dynamical channels, the merger probability depends on the total mass, rather than the mass ratio, as mass segregation and dynamical interactions may favor binaries with larger total masses [198, 204]. [154] proposed measuring the pairing function's dependence on the total mass to discriminate between formation channels, as the pairing function is expected to scale as $M_{\rm tot}^{\alpha}$ with $\alpha = 4$ in the dynamical channel modeled by [198] and $\alpha \sim 1$ for merging primordial BHs. Constraining the BBH pairing function with GW observations allows us to test these different predictions.

The pairing function has been previously studied in the context of the initial mass function for binary stars, where the degree of correlation between component stars (and the dependence on the orbital separation) remains an open question [207, 158, 163, 192, 164]. It is possible that studying the pairing function for merging BBHs may shed light on the masses of their stellar progenitors, although the relationship between a BH's mass and its progenitor star's zero-age main-sequence (ZAMS) mass is complicated by the many stages of evolution undergone by BBHs.

In the stellar context, it has been pointed out that different pairing functions affect the 1-dimensional distribution of mass ratios as well as the 1-dimensional distributions of primary and secondary masses [158]. Because the primary (secondary) mass is defined to be the more (less) massive component in the binary, even randomly drawing two components from the same underlying distribution results in the primary and secondary masses having

different distributions. Random draws can also result in very different mass ratio distributions, depending on the shape of the underlying mass distribution. We emphasize that the pairing mechanism cannot be determined by examining any one of these one-dimensional distributions independently. For example, a mass ratio distribution that favors near-unity mass ratios may simply indicate that the underlying BH mass distribution peaks in a narrow mass range, rather than that similar component masses are more likely to partner and merge. It is therefore important to examine the two-dimensional mass distribution in order to analyze whether or not there is a preference for similar-mass components.

This paper explores the BBH pairing function by analyzing the first ten LIGO/Virgo BBHs according to the mass models described in Section 5.3. The results of the analysis and implications for future detections are found in Section 5.5. In Section 5.6.2 we demonstrate the analysis on mock GW data and forecast the constraints that will be possible with ~ 50–100 more BBH detections (to be expected at the end of O3 or shortly after the start of O4). We conclude in Section 5.7. Section 5.4 describes the details of the hierarchical Bayesian analysis.

5.3 Mass Distribution Models

In the simplest case, we consider a model in which the component masses in a BBH system are independently drawn from the same underlying power-law distribution:

$$p(m) \propto m^{\gamma}, \quad m_{\rm min} < m < m_{\rm max}, \tag{5.1}$$

where γ is the power-law slope, and $m_{\rm min}$ and $m_{\rm max}$ are the minimum and maximum mass. We refer to this as the "random pairing" mass distribution [158]. We note that in this case, the marginal distributions of the primary and secondary masses are not identical, because the primary (secondary) is defined as the more massive (less massive) component. Defining

m_1 as the primary mass and m_2 as the secondary mass, the random pairing power-law distribution takes the form:

$$p(m_1, m_2 \mid \gamma, m_{\rm min}, m_{\rm max}) = \begin{cases} \frac{2(\gamma+1)^2}{\left(m_{\rm max}^{\gamma+1} - m_{\rm min}^{\gamma+1}\right)^2} m_1^{\gamma} m_2^{\gamma} & \text{if } m_{\rm min} < m_2 < m_1 < m_{\rm max} \\ 0 & \text{else.} \end{cases} \tag{5.2}$$

This implies that the primary masses follow the distribution:

$$p(m_1 \mid \gamma, m_{\rm min}, m_{\rm max}) = \frac{2\left(\gamma+1\right) m_1^{\gamma} \left(m_1^{\gamma+1} - m_{\rm min}^{\gamma+1}\right)}{\left(m_{\rm max}^{\gamma+1} - m_{\rm min}^{\gamma+1}\right)^2}, \tag{5.3}$$

while the secondaries follow:

$$p(m_2 \mid \gamma, m_{\rm min}, m_{\rm max}) = \frac{2\left(\gamma+1\right) m_2^{\gamma} \left(m_{\rm max}^{\gamma+1} - m_2^{\gamma+1}\right)}{\left(m_{\rm max}^{\gamma+1} - m_{\rm min}^{\gamma+1}\right)^2}. \tag{5.4}$$

We reiterate that the distributions in Eqs. 5.3 and 5.4 are not the same as the underlying distribution (Eq. 5.1), even though both masses are separately drawn from this distribution. In particular, the primary mass distribution will tend to favor larger masses compared to the secondary mass distribution. Furthermore, different choices of the underlying power-law parameters (γ, $m_{\rm min}$, and $m_{\rm max}$) will lead to different distributions in the mass ratio $q \equiv m_2/m_1 \leq 1$. If the underlying power-law is steep enough in either direction, mass ratios close to unity will be favored even if the two components are randomly paired.

In order to explore the pairing of two component BHs, we consider mass distributions that contain the random pairing distribution as a sub-model, but allow for deviations parameterized by a pairing function, f_p. Motivated by population synthesis models, we consider two pairing functions: one that depends on the mass ratio, q, where $q = m_2/m_1 \leq 1$, and one that depends on the total mass, $M_{\rm tot} = m_1 + m_2$. We also consider the possibility that

the probability of two BHs forming a binary and merging depends on both q and $M_{\rm tot}$.

If each BH mass is drawn from an underlying power-law distribution with slope γ, and the probability of two masses belonging to a merging binary is given by $f_p(q, M_{\rm tot} \mid \vec{\beta})$, where $\vec{\beta}$ denotes the hyper-parameter(s) of the pairing function, the mass distribution of merging BHs follows:

$$p(m_1, m_2 \mid \gamma, m_{\rm min}, m_{\rm max}, \vec{\beta}) \propto \begin{cases} m_1^{\gamma} m_2^{\gamma} f_p(\frac{m_2}{m_1}, m_1 + m_2 \mid \vec{\beta}) & \text{if } m_{\rm min} < m_2 < m_1 < m_{\rm max} \\ 0 & \text{else.} \end{cases} \tag{5.5}$$

We first consider pairing functions, f_p, that depend purely on the mass ratio. As a simple model, we assume that the pairing function follows a power-law in mass ratio with slope β_q, with a minimum mass ratio threshold required for merger, $q_{\rm min}$. The parameter $q_{\rm min}$ allows us to explore the scenario in which mergers only take place between equal component masses ($q_{\rm min} \to 1$). In this model:

$$f_p(q, M_{\rm tot} \mid \vec{\beta} = \{\beta_q, q_{\rm min}\}) \propto \begin{cases} q^{\beta_q} & \text{if } q > q_{\rm min} \\ 0 & \text{else.} \end{cases} \tag{5.6}$$

This model reduces to the random pairing model for $\left(\beta_q, q_{\rm min}\right) = \left(0, \frac{m_{\rm min}}{m_{\rm max}}\right)$.

Second, we consider pairing functions f_p that depend on the total mass of the system. A simple model in this case, motivated by the predictions of the dynamical channel of [198] and the primordial BH channel of [154], is a simple power-law in $M_{\rm tot}$:

$$f_p(q, M_{\rm tot} \mid \vec{\beta} = \{\beta_M\}) \propto M_{\rm tot}^{\beta_M}. \tag{5.7}$$

More generally, we may consider a pairing function that depends on both the mass ratio and

the total mass:

$$f_p(q, M_{\rm tot} \mid \vec{\beta} = \{\beta_q, q_{\rm min}, \beta_M\}) \propto \begin{cases} q^{\beta_q} M_{\rm tot}^{\beta_M} & \text{if } q > q_{\rm min} \\ 0 & \text{else.} \end{cases} \tag{5.8}$$

We highlight that because we consider models that reduce to random pairing under some choice of parameters, our assumed parameterization differs from the mass distribution models analyzed in [19]. The basic power-law model in [19] is defined such that the marginal $p(m_1)$ distribution follows a power-law, so that the joint mass distribution takes the form [note we have redefined α from 19, as $-\alpha$]:

$$\begin{aligned} &p(m_1, m_2 \mid \alpha, \beta, m_{\rm min} m_{\rm max}) = \\ &\qquad (\alpha+1)(\beta+1)\frac{m_1^{\alpha}}{m_{\rm max}^{\alpha+1} - m_{\rm min}^{\alpha+1}}\frac{m_2^{\beta}}{m_1^{\beta+1} - m_{\rm min}^{\beta+1}}. \end{aligned} \tag{5.9}$$

On the other hand, for the parameterizations we consider in this work, the marginal distribution of primary masses does not follow an exact power-law; instead, these parameterizations allow for the possibility that both masses in a binary are drawn from the same underlying power-law distribution.

Following the methods laid out in Section 5.4, we fit the pairing models discussed above to the first ten BBH detections in Section 5.5. As part of our results, we quantify the evidence against the random pairing and the total-mass dependent pairing models, and find that the mass-ratio dependent pairing model provides the best fit to the data.

5.4 Methods

We carry out a hierarchical Bayesian analysis to fit the hyper-parameters for each of the mass models discussed in Section 5.3. We fit only for the distribution of primary and secondary

masses, and fix the distributions of all other BBH intrinsic and extrinsic source parameters. We fix the underlying redshift distribution to follow a merger rate that is uniform in comoving volume and source-frame time. We assume that the underlying population is isotropic on the sky, with isotropic inclination angles. For definiteness we fix the spin distribution of both binary components to be uniform in spin magnitude and isotropic in spin tilt. Although this distribution is not necessarily favored by the data, the correlation between the inferred spin distribution and the inferred mass distribution is negligible, as shown in [19], which fit simultaneously for the mass and spin distribution. In particular, despite using a different spin model, we recover the results of [19] under the same mass model.

The likelihood is given by the inhomogeneous Poisson process likelihood [169, 179, 19]. For $N_{\rm obs}$ independent events, the likelihood of the data d given hyper-parameters θ is:

$$p(d \mid \theta) \propto e^{-\mu(\theta)} \prod_{i=1}^{N_{\rm obs}} \int p(d_i|m_1, m_2) \, \frac{\mathrm{d}\mathcal{R}}{\mathrm{d}m_1 m_2}(\theta) \, \mathrm{d}m_1 m_2, \tag{5.10}$$

where $p(d_i \mid m_1, m_2)$ denotes the likelihood of an individual event's data given its component masses, $\frac{\mathrm{d}\mathcal{R}}{\mathrm{d}m_1 m_2}(\theta)$ is the differential merger rate density, which integrates to the total merger rate density $\mathcal{R}$ and is given by $\mathcal{R}p(m_1, m_2 \mid \theta)$, and $\mu(\theta) = \mathcal{R}\langle VT\rangle_\theta$ denotes the expected number of detections given $\mathcal{R}$ and the sensitive spacetime volume $\langle VT\rangle_\theta$ of the detector network to a given population of BBHs with hyper-parameters θ.

We assume that the merger rate density is uniform in comoving volume and source-frame time, and calculate $\langle VT\rangle_\theta$ according to a semi-analytic prescription [103, 102]. Following [19], we assume that a single-detector signal-to-noise (SNR) threshold of 8 is necessary and sufficient for detection, and that the detector's noise curve is described by the Early High Sensitivity power spectral density (PSD) for advanced LIGO [7]. The validity of these assumptions is discussed in [19]. Unlike in [19], in this work we do not calibrate the $VT(m_1, m_2)$ to the results of injection campaigns into the detection pipelines. As we demonstrate by ex-

plicitly comparing our results to those of [19] in Section 5.5, using the uncalibrated VT calculation leads to a slight bias in our inference of the overall-merger rate, with the median shifting by a factor of ~ 1.7, as expected from Fig. 9 in [19]. However, this does not affect the inferred shape of the mass distribution, which is our primary interest in this work. We also neglect the effect of non-zero spins in the estimation of VT, as spins have a sub-dominant effect on the sensitivity [19], especially given that existing detections disfavor a significant population of highly spinning systems.

Note that if we marginalize over the rate density $\mathcal{R}$ with a flat-in-log prior, the likelihood takes the form [109, 179]:

$$p(d \mid \theta) \propto \prod_{i=1}^{N_{\rm obs}} \frac{\int p(d_i \mid m_1, m_2) p(m_1, m_2 \mid \theta)\, \mathrm{d}m_1 \mathrm{d}m_2}{\int VT(m_1, m_2) p(m_1, m_2 \mid \theta)\, \mathrm{d}m_1 \mathrm{d}m_2}. \tag{5.11}$$

To get the individual-event likelihood term $p(d_i \mid m_1, m_2)$ that appears in Eq. 5.10, we use the publicly available `IMRPhenomPv2` posterior samples for the ten BBH detections in O1 and O2 [123]. There is a negligible difference between the mass posteriors derived with the `IMRPhenomPv2` waveform [140, 148] and the `SEOBNRv3` waveform [202]. The individual-event posteriors were calculated under priors that are flat in *detector*-frame masses, and "volumetric" in luminosity distance, d_L. In terms of source-frame masses and cosmological redshift z, the default event-level prior is therefore [19]:

$$p(m_1, m_2, z) \propto d_L(z)^2 (1+z)^2 \left(d_C(z) + \frac{(1+z) d_H}{E(z)} \right), \tag{5.12}$$

where d_C is the comoving distance and $d_H = c/H_0$ is the Hubble distance, and $E(z) = H(z)/H_0$ [133]. We divide out by these priors in our analysis to get a term that is proportional to the likelihood rather than the posterior. We fix the cosmological parameters to the best-fit Planck 2015 values [24] throughout for consistency with [19] and [17].

We sample from the overall likelihood of Eq. 5.10 using PyMC3 [217]. In all models

considered, we choose priors that are flat over $m_{\rm min}$, $m_{\rm max}$ and the power-law slope γ within the ranges $3\ M_\odot < m_{\rm min} < 10\ M_\odot$, $35\ M_\odot < m_{\rm max} < 100\ M_\odot$ and $-4 < \gamma < 2$. We take a flat-in-log prior on the rate $p(\mathcal{R}) \propto 1/\mathcal{R}$. Unless they are fixed to some value, we take a flat prior on β_q in the range $-4 < \beta_q < 12$ and a flat prior on β_M in the range $0 < \beta_M < 12$. Because the prior range of the minimum mass ratio $q_{\rm min}$ depends on two other free parameters, $m_{\rm min}$ and $m_{\rm max}$, we introduce another parameter $q_{\rm scale}$, defined so that:

$$q_{\rm min} = m_{\rm min}/m_{\rm max} + q_{\rm scale}(0.95 - m_{\rm min}/m_{\rm max}), \tag{5.13}$$

and, unless it is fixed, we sample over $q_{\rm scale}$ with a flat prior from 0 to 1, so that $m_{\rm min}/m_{\rm max} < q_{\rm min} < 0.95$. We restrict the upper limit of $q_{\rm min}$ to slightly below 1 in order to avoid sampling issues, as the mass ratio of any individual GW event is measured with a finite resolution, and this prevents $q_{\rm min}$ from being resolved arbitrarily close to $q_{\rm min} = 1$. When using Model B from [19], we use their same priors, with the exception of lower prior boundary for $m_{\rm min}$, which we take to be $3\ M_\odot$ rather than $5\ M_\odot$.

For those models which contain random-pairing as a subset, we quantify the evidence for the random-pairing hypothesis versus the full model by calculating the Savage-Dickey density ratio (SDDR), which is defined as the ratio of the posterior probability to the prior probability at the given point in parameter space [84]. We calculate the evidence between the pure mass-ratio dependent pairing function of Eq. 5.6 and the pure total-mass dependent pairing function of Eq. 5.7 by introducing a more general model that contains both models of interest as nested models. This general model contains a mixture parameter x, where x denotes the amplitude of the mass-ratio dependent component and $(1 - x)$ denotes the amplitude of the total-mass dependent component. By sampling from this mixture model, the recovered likelihood at $x = 1$ compared to $x = 0$ denotes the evidence in favor of a pure mass-ratio dependent pairing function.

5.5 Results

5.5.1 LVC Model

We begin by recovering the results of [19] under the same mass model given by Eq. 5.9, equivalent to Model B of [19], in order to demonstrate consistency between our methods. Although we use slightly different assumptions regarding the spin distribution and the selection effects calculation (see Section 5.4), we recover nearly-identical posterior distributions on the population hyper-parameters: $\alpha = -1.7^{+1.8}_{-1.5}$, $\beta = 6.1^{+5.4}_{-5.7}$, $m_{\rm min} = 7.3^{+1.7}_{-3.7}\ M_\odot$, and $m_{\rm max} = 41.8^{+18.6}_{-5.5}\ M_\odot$. This is to be compared with $\alpha = -1.6^{+1.7}_{-1.5}$, $\beta = 6.7^{+4.8}_{-5.9}$, $m_{\rm min} = 7.9^{+1.2}_{-2.5}\ M_\odot$, and $m_{\rm max} = 42.0^{+15.0}_{-5.7}\ M_\odot$ found by [19]. As described in Section 5.4, our prior on $m_{\rm min}$ starts at 3 $M_\odot$ rather than 5 $M_\odot$. Furthermore, recall that our convention for the power-law slope α has a sign flip compared to the convention in [19].

With the current set of events, the data cannot distinguish between the mass model of [19] (Eq. 5.9) and the models we consider in this work, and they all give consistent results for the inferred mass distribution $p(m_1, m_2)$. However, the parameters of our models have a different interpretation from the mass model of [19]. While the power-law slope α of Eq. 5.9 refers to the power-law of the primary mass distribution, the power-law slope γ of this work refers to the underlying mass distribution power-law from which both primary and secondary BHs are drawn. The additional parameters β_q, $q_{\rm min}$ and β_M in our models allow us to explore whether the pairing between the two component masses is random or whether (and how) it depends on the mass ratio or total mass of the system, according to Eq. 5.8).

5.5.2 Random Pairing

Figure 5.1 shows the results of fitting the random-pairing model (Eq. 5.2 with free parameters γ, $m_{\rm min}$, $m_{\rm max}$), the mass ratio power-law model (Eq. 5.6 with free parameters γ, $m_{\rm min}$, $m_{\rm max}$, β_q) and the total mass power-law (Eq. 5.7 with free parameters γ, $m_{\rm min}$, $m_{\rm max}$, β_M)

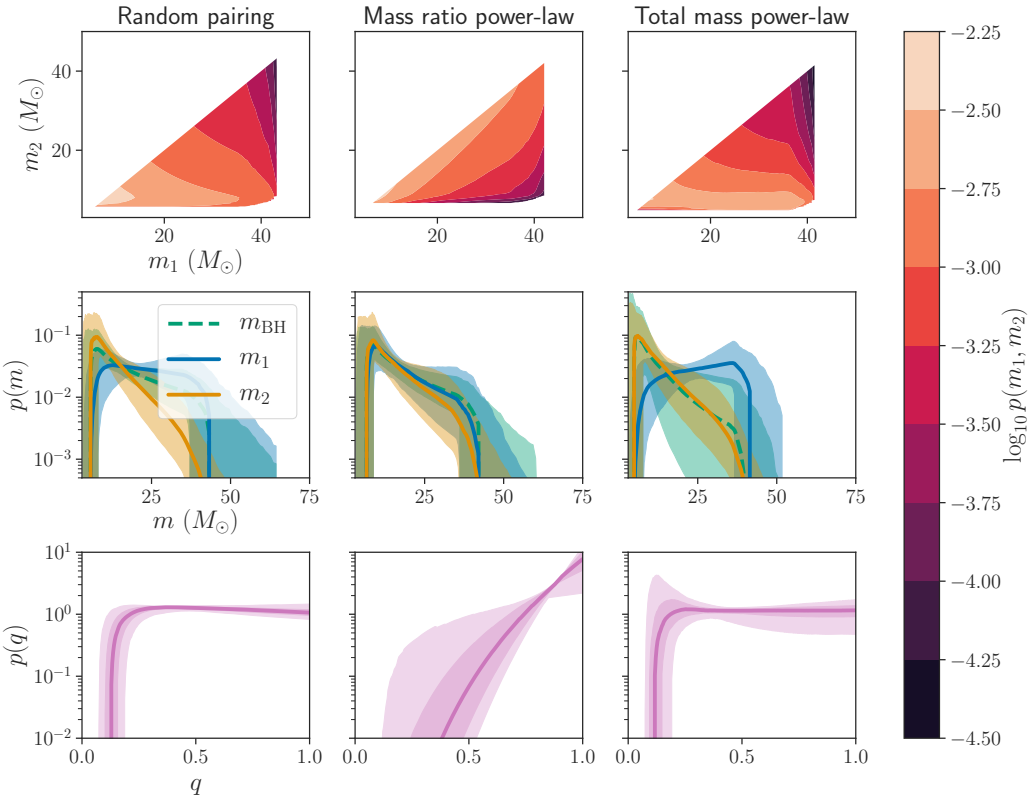

Figure 5.1: *Top row:* **Joint m_1–m_2 distribution as inferred from the ten BBHs** assuming a mass distribution given by Eq. 5.8 with free parameters γ, $m_{\rm min}$, $m_{\rm max}$ (left column), γ, $m_{\rm min}$, $m_{\rm max}$ and β_q (middle column), and γ, $m_{\rm min}$, $m_{\rm max}$ and β_M (right column). In each case, those parameters that are not free are fixed to $\beta_q = \beta_M = 0$ and $q_{\rm min} = m_{\rm min}/m_{\rm max}$. The color scale indicates the median $\log_{10}$ of the merger rate density as a function of the two masses. *Middle row:* Marginal distributions of single BH masses (green), along with the primary masses (blue) and secondary masses (yellow) of component BHs in binary systems. These distributions are inferred by fitting the ten BBH detections to the model of the corresponding column. The line shows the median merger rate density as a function of mass, while the shaded bands show symmetric 90% credible intervals. *Bottom row:* Marginal distribution of the mass ratio implied by the fits to the three models. The solid line and dark (light) bands denote median and 50% (90%) credible intervals on the merger rate as a function of mass ratio.

to the ten BBHs from the first two observing runs. We use flat priors on all free parameters, with uninformative prior bounds listed in Section 5.4. In each case, the parameters of Eq. 5.8 that are not left free are fixed to the default values $\beta_q = \beta_M = 0$, $q_{\rm min} = m_{\rm min}/m_{\rm max}$.

If we fix the pairing to be random, we find $\gamma = -1.1^{+1.0}_{-0.9}$. However, as shown in the left-hand, middle panel in Figure 5.1, and explained in Section 5.3, this does not imply that the one-dimensional marginal distributions of the primary and secondary masses follow this common power-law; the primary masses follow a flatter distribution while the secondary masses follow a steeper distribution. Note that the inferred mass ratio distribution (bottom row, left-hand panel) in this case is inferred to be nearly flat across the range ~ 0.15–1. This is a consequence of this particular fit to the random-pairing model; in general the marginal mass ratio distribution can slope significantly upwards depending on the value of γ. The lower-limit on the mass ratio in the random-pairing model is given by the ratio $m_{\rm min}/m_{\rm max}$, and is constrained to ~ 0.15 in this case due to the measurements $m_{\rm min} \sim 7\ M_\odot$ and $m_{\rm max} \sim 40\ M_\odot$.

5.5.3 Mass Ratio Dependent Pairing

The effect of introducing β_q as a free parameter, while still fixing $q_{\rm min} = m_{\rm min}/m_{\rm max}$ and $\beta_M = 0$, is shown the middle column of Figure 5.1. Under this model extension, the data display a clear preference for mass ratios close to unity (bottom row, middle column), which implies more overlap between the primary and secondary mass distributions (middle row, central panel). We infer $\beta_q = 7.0^{+4.5}_{-5.5}$, and find that $\beta_q \leq 0$ is ruled out with probability 0.987. This suggests that *the random pairing model ($\beta_q = 0$) is strongly disfavored by the data.* Meanwhile, the underlying mass distribution power-law slope is inferred to be a bit steeper than in the random-pairing case, with $\gamma = -1.3^{+0.9}_{-0.8}$ compared to $\gamma = -1.1^{+1.0}_{-0.9}$ for the random-pairing model. All models essentially agree on $m_{\rm min} \sim 6.7^{+1.9}_{-3.2}\ M_\odot$ and $m_{\rm max} \sim 41.9^{+18.2}_{-5.7}\ M_\odot$.

We fit the model with both β_q and $q_{\rm min}$ left free in Figure 5.2, which displays the posterior distributions on the five hyper-parameters (three parameters to characterize the 1-dimensional mass distribution, and two – β_q and $q_{\rm min}$ – to characterize the pairing function) as a corner plot [111]. We find a strong preference for near-equal mass ratios, inferring that 99% of merging BBHs have mass ratios between $q_{1\%} = 0.66^{+0.25}_{-0.28}$ and unity. For reference, under Model B from [19], we find $q_{1\%} = 0.59^{+0.13}_{-0.33}$. Meanwhile, we find that the median mass ratio is $q_{50\%} = 0.91^{+0.05}_{-0.17}$. This agrees very closely with the findings of [215], who find a preference for population distributions with an average mass ratio $\bar{q} = 0.89^{+0.08}_{-0.18}$. In fact, the current set of detections is consistent with all binaries consisting of equal component masses (to within 5%, with $q_{\rm min} = 0.95$; the maximum value permitted by our prior). We find that $q_{\rm min} = 0.95$ is five times more likely than $(\beta_q, q_{\rm min}) = (0, m_{\rm min}/m_{\rm max})$, meaning that it is *five times more likely that all binaries consist of equal component masses than that they are randomly paired*. Although we do not include the events of [248, 247] or [196] in our analysis (in order to avoid assuming an incorrect selection function and biasing our results), we note that all of their detections are also consistent with mass ratios of unity. This suggests an even stronger preference for near-equal component masses in the underlying population.

A strong preference for near-equal component masses with $q_{\rm min} \gtrsim 0.9$ is consistent with the chemically homogeneous/massive overcontact binary evolutionary channel [186]. However, our results are also consistent with a milder preference for near-unity mass ratios, which may be expected from classical isolated binary evolution or dynamical formation [177].

5.5.4 Total Mass Dependent Pairing

The effect of introducing β_M as a free parameter, while fixing $q_{\rm min} = m_{\rm min}/m_{\rm max}$ and $\beta_q = 0$, is shown in the right-hand column of Figure 5.1. We note that these results are similar to the random pairing case, as there is a strong degeneracy between the isolated BH mass power-law slope γ and the pairing function power-law slope β_M. We do not recover

significant constraints beyond the linear combination $2\gamma + \beta_M \approx -1.1^{+1.0}_{-0.9}$ (matching the constraints reported in Section 5.5.2 for the case $\beta_M = 0$). The degeneracy between β_M and γ causes the constraints on the isolated BH mass function (shown in green in the left-hand column, middle row of Figure 5.1) to be degraded. In addition to the degraded constraints on γ, the constraints on $m_{\rm min}$ are also less informative, as there is a significant correlation between $m_{\rm min}$ and γ, with shallow (less negative) power-law slopes corresponding to smaller allowed values of $m_{\rm min}$. However, $m_{\rm max}$ remains very well-measured at $m_{\rm max} \sim 41.9$.

It is interesting to note the inferred value of the power-law slope γ for the cases $\beta_M = 4$ (predicted by some globular cluster simulations) and $\beta_M = 1$ (predicted by primordial BH channel). For $\beta_M = 4$, we find $\gamma \sim -2.5$, which is close to the Salpeter initial mass function for massive stars. The interpretation in this case is that BHs in globular clusters, where the merger probability roughly follows $M_{\rm tot}^4$ [199], are distributed according to a power-law mass distribution with slope $\gamma \sim -2.5$. This agrees with the findings of [204], who forward-model dynamical mergers for various initial BH mass functions and compare the model predictions to the LIGO/Virgo data to infer an initial BH mass function with slope $-2.35^{+0.36}_{-0.55}$ (68% credibility). Meanwhile, $\beta_M = 1$ implies a shallower underlying mass distribution for single BHs, with $\gamma \sim -1$.

If we leave all three parameters of the pairing function, $q_{\rm min}$, β_q, and β_M, free, we recover identical constraints on $q_{\rm min}$ and β_q as we do in the $\beta_M = 0$ case. The additional freedom in β_M is fully absorbed by the degeneracy with γ, and we recover the prior on β_M. This is expected, because the data prefers equal mass systems with $m_1 = m_2$, and in this limit, Eq. 5.8 reduces to:

$$p(m_1, m_2) \propto m_1^{2\gamma - \beta_M}, \text{ for } m_1 = m_2 \tag{5.14}$$

yielding a complete degeneracy along $2\gamma - \beta_M = \text{const}$.

5.5.5 Comparison of Pairing Functions

The data strongly prefers a mass-ratio dependent pairing function that favors mergers between similar-mass components over random pairing, as discussed in Section 5.5.3. However, unless one has a strong prior on the single BH mass distribution that would push β_M away from zero (e.g. a prior belief that $\gamma < -2$ favors $\beta_M > 0$), there is no evidence that the total mass plays a role in the pairing function. This is apparent by the fact that when we fit Eq. 5.8 to the data with all six hyper-parameters free, we recover the prior on β_M. Computing the evidence ratio in favor of a model in which $\beta_M = 0$ to one in which $(\beta_q, q_{\rm min}) = (0, m_{\rm min}/m_{\rm max})$, we find that it is ~ 6 times more likely that the pairing function depends on the mass ratio than on the total mass.

5.5.6 Posterior predictive distributions

In the following, because there is no evidence that the data prefers nonzero β_M, we fix $\beta_M = 0$ in the population model of Eq. 5.8 (reducing to the model of Eq. 5.6). Using the recovered posteriors on the population hyper-parameters, shown as a corner plot in Figure 5.2, we calculate the posterior population distribution $p(m_1, m_2 \mid \text{data})$, shown in Figure 5.3. We define the posterior population distribution as in [19]; this refers to the distribution of true mass values marginalized over the hyper-parameter posteriors for a given population model:

$$p(m_1, m_2 \mid d) = \int p(m_1, m_2 \mid \theta) p(\theta \mid d) \mathrm{d}\theta. \tag{5.15}$$

Here, $p(\theta \mid d)$ refers to the posterior distribution on the population model's hyper-parameters inferred from the ten BBH events. Figure 5.3 shows draws from the posterior population distribution on the true masses of the underlying BBH population (blue) as well as the true masses of *detected* systems (orange), found by applying selection effects to the underlying mass distribution. For comparison, the mass measurements of the O1 and O2 detections are

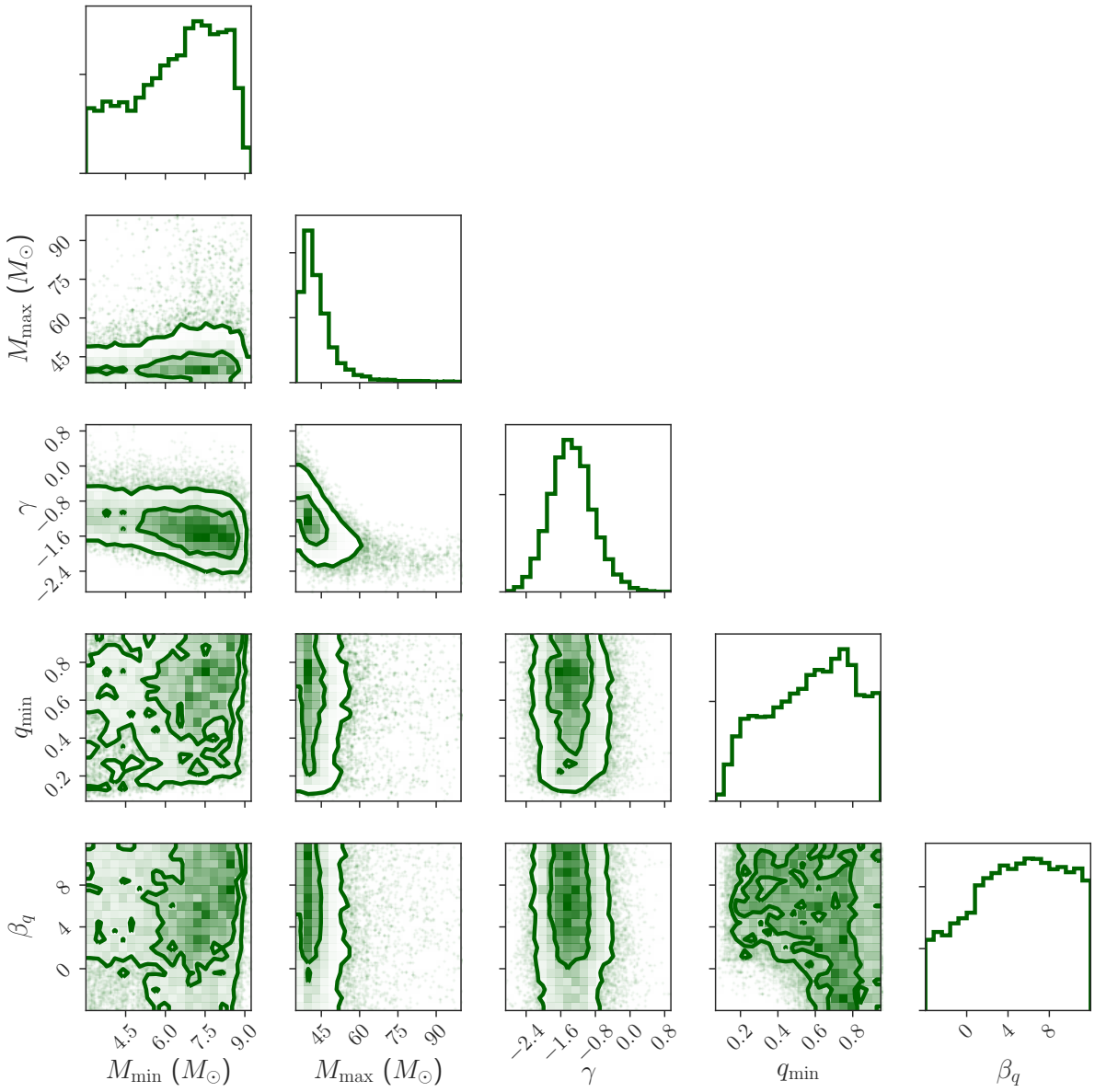

Figure 5.2: **Posterior on the hyper-parameters of the power-law model with the mass-ratio dependent pairing** (Eq. 5.6) fit to the ten BBH detections from O1 and O2. In the two-dimensional plots, the contours denote 50% and 90% posterior credible regions.

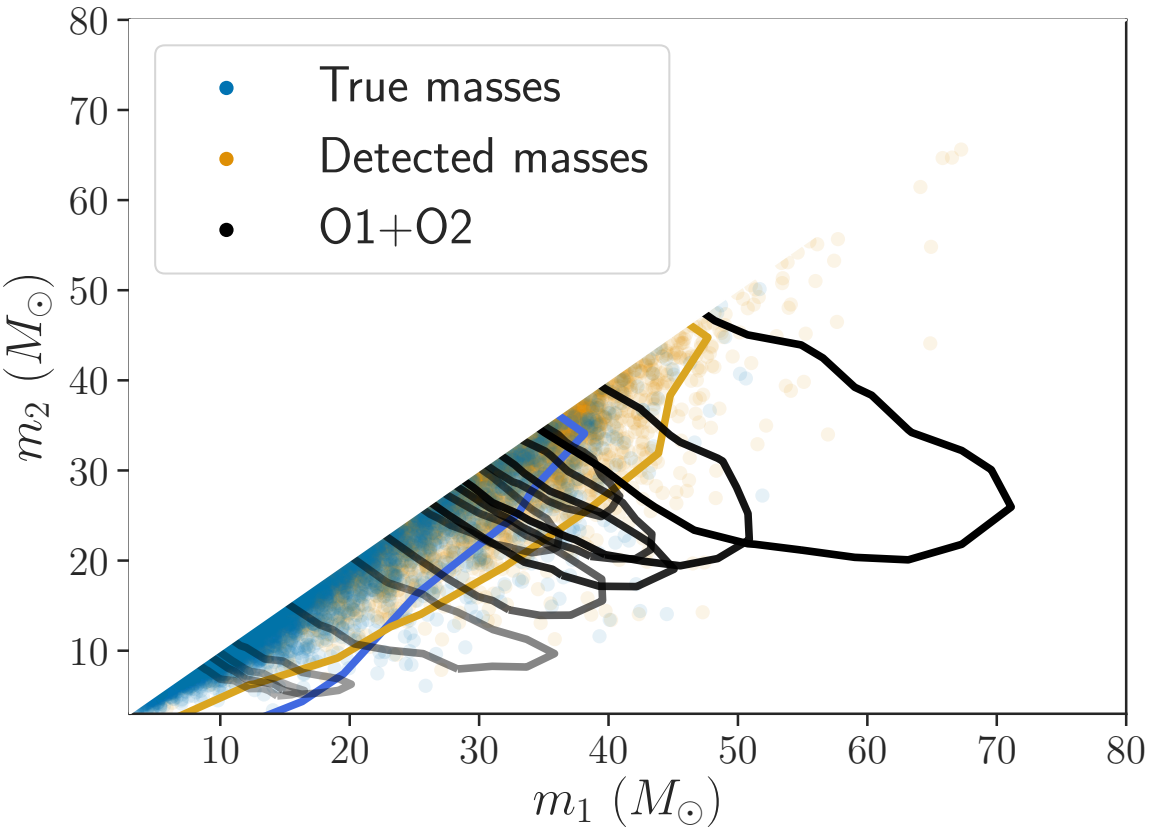

Figure 5.3: **Posterior population distribution of the component masses in BBH binaries**, as inferred from the mass-ratio dependent pairing model. The true masses of the *underlying* population are represented by the blue points and 90% credible region, while the orange points represent the *detected* population, accounting for selection effects that favor more massive systems. In grayscale are the mass measurements of the ten LIGO/Virgo O1 and O2 detections. The contours denote 90% credible intervals. All detected systems are consistent with equal component masses $m_1 = m_2$.

shown in black.

Using the posterior population distribution on the masses m_1 and m_2, we calculate the corresponding distribution on the mass ratio q in Figure 5.4. The dashed blue line in Figure 5.4 shows the mass ratio distribution in the underlying population, corresponding to the blue m_1–m_2 distribution of Figure 5.3, while the orange line shows the mass ratio distribution among detected systems, corresponding to the orange m_1–m_2 distribution of Figure 5.3. While selection effects have a significant effect on the two-dimensional m_1–m_2 distribution, the 1-dimensional *mass ratio* distribution among detected systems is nearly identical to the mass ratio distribution in the underlying population.

We expect 90% of detected events to have their true masses fall within the orange credible region of Figure 5.3. In terms of the mass ratio distribution, we expect that 90% of detected events will have mass ratios $q > 0.73$, and 99% of detected events will have mass ratios $q > 0.51$. We can take these predictions one step further by simulating the *measured* mass ratio values for detected events, which accounts for measurement uncertainty in addition to the selection effects. Given the true masses of a detected event, we generate a mock posterior to represent how those masses would be measured in LIGO/Virgo data. The mock posteriors are generated according to the prescription described in Appendix 5.6.1. We summarize the expected mass ratio posteriors from anticipated detections as the green dashed line (median) and shaded band (symmetric 90% interval) in Figure 5.4. We refer to this green band as the "posterior predictive process." Based on the first ten detections, and assuming that all detections are described by the same population model assumed here, we expect that 90% of future detections will have recovered mass ratio posteriors that lie within the shaded band. We see that measurement uncertainty plays a significant role in shifting the observed mass ratio posteriors (with the default flat-in-component-mass priors) away from 1 relative to $q_{\rm true}$.

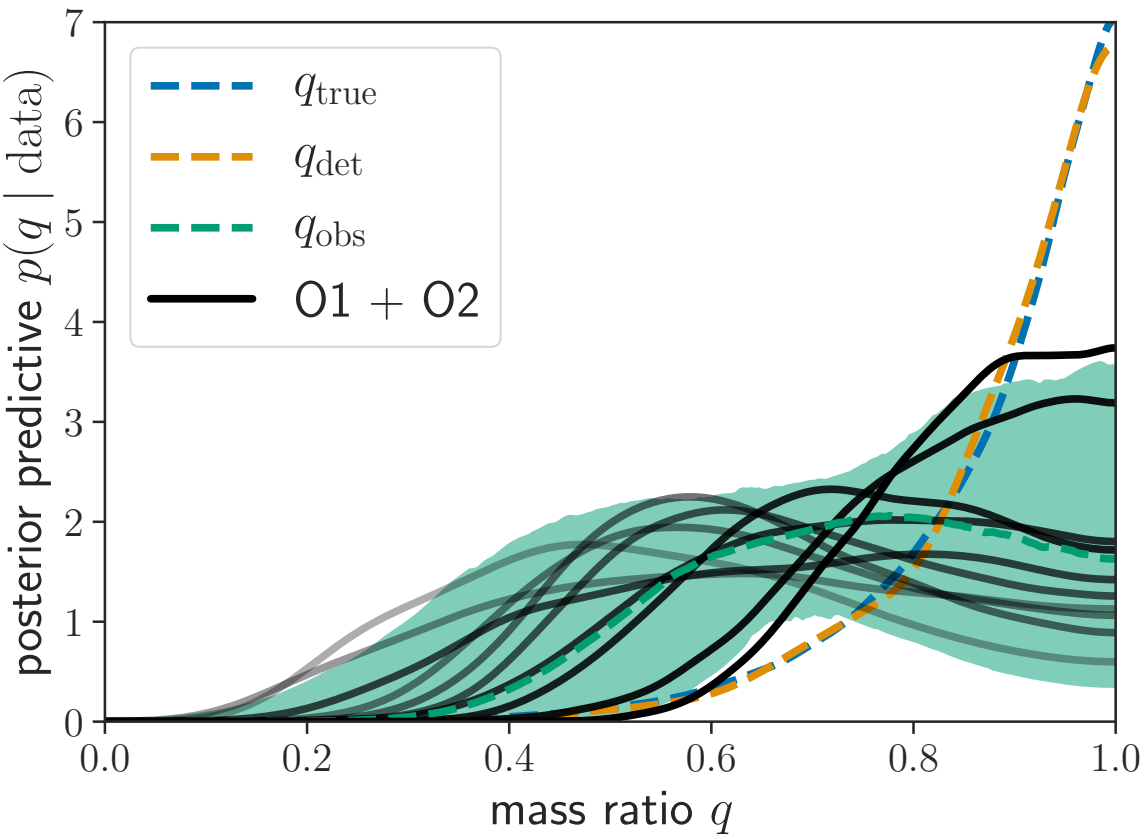

Figure 5.4: **Posterior population distribution of the mass ratio** q_{true} in the *underlying* population (dashed blue line), the mass ratio q_{det} among *detected* systems (dashed orange line), and the posterior predictive process of the *measured* mass ratio q_{obs} (dashed green line and shaded band), accounting for detection efficiency and measurement uncertainty. These distributions are inferred by fitting the ten BBHs from O1 and O2 to the mass distribution model described by Eq. 5.6. If all BBHs belong to this population, we expect that 90% of the recovered posteriors from detected BBHs will fall within the shaded green region. The grayscale lines show the posterior probability distributions of the ten observed BBHs. Note that measurement uncertainty shifts the posteriors on the mass ratio for individual systems to smaller values relative to the true mass ratio.

5.6 Simulations

We expect to have tens more BBH detections by the end of LIGO/Virgo's third observing run in mid-2020, and hundreds of detections within a few more years [7]. In this section, we explore the expected mass distribution constraints from tens to hundreds of detections under the models considered here. We perform our analysis on mock GW detections that we generate from known underlying distributions. We follow a simplified yet realistic method for generating mock measurements from the underlying population and ensure that the mock primary and secondary masses are measured with uncertainties typical to second-generation GW detectors [251]. The method for generating mock detections is described in subsection 5.6.1 and the results are described in subsetction 5.6.2.

5.6.1 Method for Simulating Detections

In generating mock detections, we assume that the underlying population follows a uniform in comoving volume and source-frame time merger rate, with isotropic sky positions and inclinations, and zero spins. The true component masses are drawn from the given population distribution. We note that the assumptions of fixed redshift and spin distributions are unlikely to affect the inference of the pairing function (mass ratios are measured independently of redshift, and excluding spins did not make a difference in the O1 and O2 analysis); however, these distributions can be fit jointly with the mass distribution and marginalized over [19].

Given the true parameters of the binary, we calculate the SNR of the signal in a single detector, assuming that the noise is described by the Mid-High Sensitivity PSD as expected for O3 for the LIGO detectors [7]. We assume that the binary is then detected if it passes a single-detector SNR threshold of 8. In order to assign measured component masses to each detected binary, we assume that the fractional uncertainty on the source-frame chirp mass

follows $\frac{\sigma_{\mathcal{M}}}{\mathcal{M}} = \frac{8}{\rho}\left(0.01 + \left(\frac{0.2z}{1+z}\right)^2\right)^{1/2}$, where z is the true redshift, while the uncertainty on the symmetric mass ratio $\eta \equiv \frac{m_1 m_2}{(m_1+m_2)^2}$ follows $\sigma_\eta = 0.03\frac{8}{\rho}$, where ρ is the single-detector SNR of the source. Given a true value of $\mathcal{M}$ and η for each binary, we randomly draw $\mathcal{M}_{\rm obs}$ from a log-normal distribution centered on $\mathcal{M}$ with standard deviation $\sigma_{\mathcal{M}}$, and $\eta_{\rm obs}$ from a normal distribution centered on η with standard deviation σ_η. With these values of $\mathcal{M}_{\rm obs}$ and $\eta_{\rm obs}$ and their assumed known distributions about the true chirp mass $\mathcal{M}$ and symmetric mass ratio η, we generate mock posterior samples for the component masses m_1 and m_2 under flat priors using the Monte-Carlo sampler PyStan [60]. These uncertainties are typical of the O2 detections, and result in typical 90% measurement uncertainties on the source-frame component masses of $\approx 50\%$, with a distribution of uncertainties that matches the one in [251].

5.6.2 Simulation Results

The expected constraints from 60 detections [similar to what we expect by the end of O3; 7] are shown in Figure 5.5 for a simulated population described by Eq. 5.6 with $m_{\rm min} = 7\ M_\odot$, $m_{\rm max} = 40\ M_\odot$, $\gamma = -1$, $\beta_q = 6$, and $q_{\rm min} = m_{\rm min}/m_{\rm max} = 0.175$. We find with 60 events from this simulated population, we can typically rule out random-pairing with Bayes factors $\gtrsim 1000$. These projections are conservative because the deviations from random-pairing in the chosen mock population are not very large compared to the values of $q_{\rm min}$ and β_q that are favored by the first ten events. The parameters that govern the pairing mechanism will become increasingly well-constrained, although with large correlations between them. The simulated 60 events shown here yield $\beta_q = 3.4^{+4.6}_{-5.9}$ and $q_{\rm scale} = 0.5^{+0.4}_{-0.4}$ ($q_{\rm min} = 0.5^{+0.3}_{-0.3}$). Meanwhile, the parameters of the underlying power-law mass distribution will also become well-constrained. With 50 more events, we expect to constrain $m_{\rm max}$ to a couple of solar masses (this particular realization yields $m_{\rm max} = 40.6^{+2.4}_{-1.7}$) and the power-law slope γ to a 90% credible interval of < 1 ($\gamma = -1.1^{+0.4}_{-0.3}$). Note that if β_M is left free in the pair-

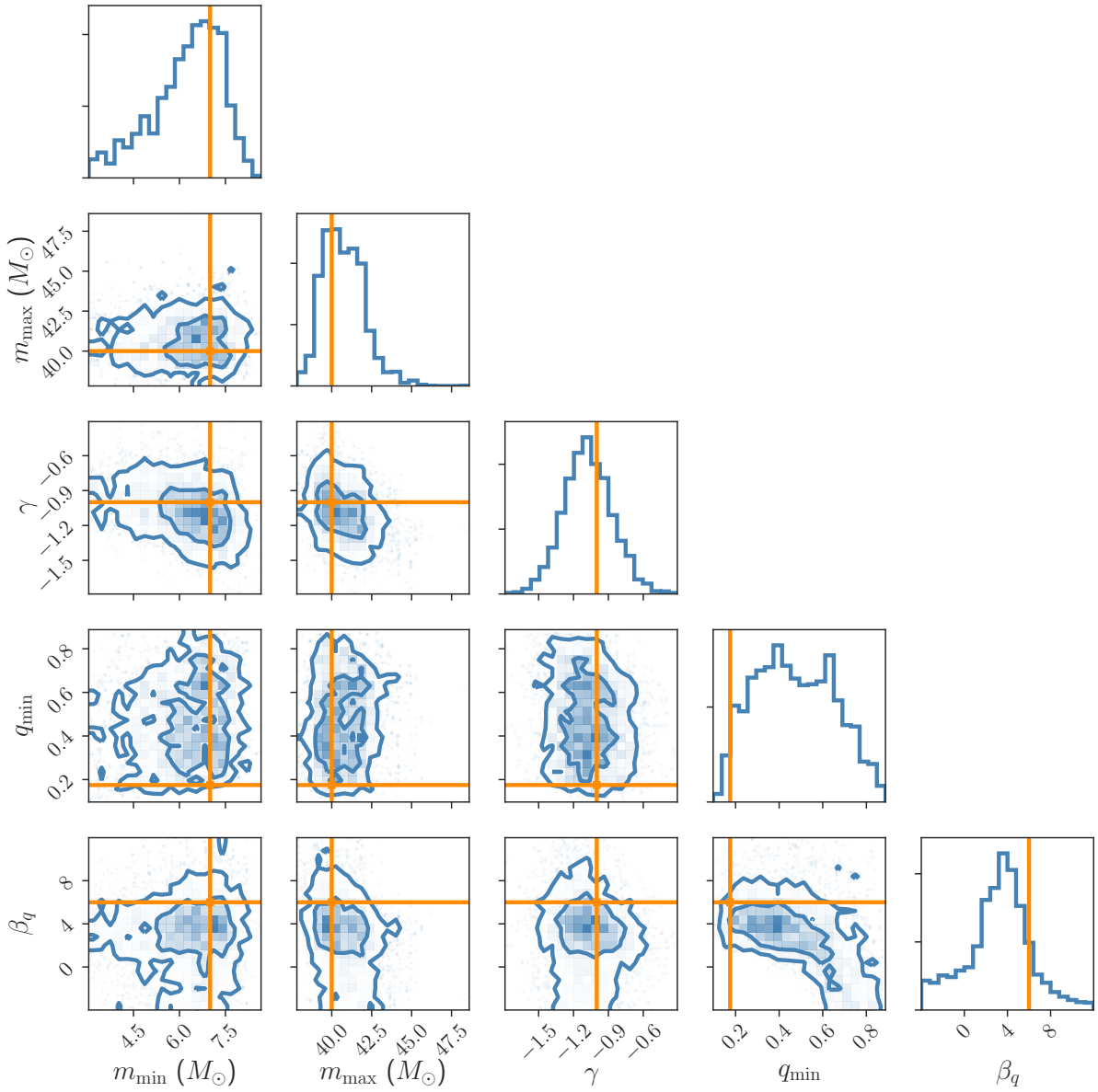

Figure 5.5: **Constraints on the population hyper-parameters for a simulated population of 60 BBH detections** that follow Eq. 5.6 with $m_{\rm min} = 7\ M_\odot$, $m_{\rm max} = 40\ M_\odot$, $\gamma = -1$, $\beta_q = 6$, and $q_{\rm min} = m_{\rm min}/m_{\rm max} = 0.175$. These injected hyper-parameter values are shown in orange lines. In the two-dimensional plots, the contours show 50% and 90% posterior credible regions.

ing function, the measurement of γ will become less informative, as we constrain a linear combination of β_M and γ (see Section 5.5). With 100 events, these constraints will improve roughly as $1/N$ and $1/\sqrt{N}$ for $m_{\rm max}$ and γ, respectively[1]: 100 simulated events gives $m_{\rm max} = 38.9^{+1.4}_{-0.9}\ M_\odot$ and $\gamma = -0.8^{+0.3}_{-0.3}$. It may take more events for the constraints on $m_{\rm min}$ to become interesting, because the detector sensitivity is a steep function of BH mass, and most detections are at the high end of the mass function [105], or beyond [107]. For a flat prior starting at $3\,M_\odot$, 60 mock events give $m_{\rm min} = 6.5^{+1.3}_{-2.4}$. However, the constraints on $m_{\rm min}$, like the constraints on γ, are less informative if β_M is allowed to vary.

5.7 Conclusion

We have fit the mass distribution of merging BBHs with a simple model that parameterizes the pairing function between the two components in a binary. We highlight the importance of comparing the full two-dimensional mass distribution of BBHs, because it is impossible to disentangle the overall BH mass distribution from the pairing function when considering only one-dimensional distributions of the primary/ secondary mass or the mass ratio.

Based on the first ten LIGO/Virgo BBH detections, we conclude that component BHs are not randomly paired in a binary; rather, the pairing likely favors components of comparable masses. We find that it is five times more likely that mergers only take place between equal (to within 5%) mass BHs than that component BHs are randomly drawn from the same underlying distribution. Our fits imply that 99% of mass ratios among the population of merging BBHs are greater (closer to equal mass) than $q_{1\%} = 0.66^{+0.25}_{-0.28}$. This is to be compared with an expected value of $q_{1\%} = 0.15^{+0.07}_{-0.06}$ for the random pairing scenario. We predict that among detected BBHs, 90% will have mass ratios $q > 0.73$, and 99% will have mass ratios $q > 0.51$. Meanwhile, we find no evidence that the pairing function depends on

1. Because $m_{\rm max}$ is a sharp feature, its measurement converges faster than the typical $1/\sqrt{N}$ [see e.g. 64, 143]. This makes it particularly useful as a feature to constrain cosmology [98].

the total mass of the system, contrary to the predictions of some dynamical and primordial BH formation channels [199, 154, 204].

The current constraints on the pairing function remain compatible with a range of formation channels, with the exception of those that favor random pairing or a preference for unequal mass ratios [190]. All binaries detected so far are consistent with equal mass components, which is compatible with predictions from the massive overcontact binary/chemically homogeneous formation channel, in which mass transfer may lead to very nearly equal mass components with minimum mass ratios $q_{\rm min} \sim 0.9$ [186]. However, the current constraints on the pairing function are also compatible with scenarios that more mildly prefer equal mass components, such as common envelope binary evolution, which tends to result in mergers with $q \gtrsim 0.5$ [177], or dynamical interactions in globular clusters, which tend to result in mergers with median mass ratios $q_{50\%} = 0.9$ [257], consistent with our measurement $q_{50\%} = 0.91^{+0.05}_{-0.17}$. On the other hand, some dynamical channels predict that the pairing function should scale with the total mass of the system. While we cannot rule this out with ten detections, it is 6 times more likely that the pairing function has some mass-ratio dependence rather than depending on total mass alone.

Although the data does not call for a total-mass dependence, it remains possible that the pairing probability depends on the total mass in addition to the mass ratio. If the merger probability scales with total mass as $M_{\rm tot}^{\beta_M}$, the implied power-law slope of the mass distribution among single BHs is roughly $\gamma \sim -\frac{1+\beta_M}{2}$. Thus, prior belief that the BH mass spectrum is steep (with $\gamma \lesssim -2.2$) would suggest that the pairing function depends on the total mass with $\beta_M > 0$.

By the end of O3, the details of the pairing function will be better constrained (compare the joint posterior on $q_{\rm min}$ and β_q in Figure 5.2—the current constraints—with Figure 5.5—the constraints we expect by the end of O3). We hope that these results will enable detailed comparisons with the predictions of the full 2-dimensional merger rate $\mathcal{R}(m_1, m_2)$ from

population synthesis simulations.

As usual, our results rely on the assumption that there is a single population of BBHs that is adequately described by our simple parameterized model [see e.g. 85]. One way to test the validity of this assumption with future detections is to compare them against the posterior predictive distribution (for example, Figures 5.3 and 5.4) inferred from the model. We conclude that the universe does not assemble its black-hole binaries at random, and future constraints of the pairing function we have introduced above will yield important insights into these formation processes.

CHAPTER 6

THE MOST MASSIVE BINARY BLACK HOLE DETECTIONS AND THE IDENTIFICATION OF POPULATION OUTLIERS

One of the most significant population features that we discovered from LIGO-Virgo's first two observing runs is a drop-off in the black hole mass spectrum at $\sim 45\ M_\odot$ (see Chapter 2). However, we realized that interpreting this population result in the context of the individual detections required some subtlety. In particular, one of the black holes we analyzed had a measured mass of $\sim 50\ M_\odot$; can this event be consistent with a population that claims a maximum mass of $45\ M_\odot$? In this chapter, we explain how to re-interpret the properties of individual events in the context of the global population under a hierarchical Bayesian statistical framework. Noise fluctuations will frequently cause us to overestimate the masses of individual black hole events, but we can reduce this uncertainty by jointly fitting the masses of individual events as well as the overall mass distribution to *all* events. Additionally, we show how to check for consistency between the individual detections and the population inference by carrying out several posterior predictive checks to quantify the "goodness-of-fit" of a model to the data. Finally, we describe techniques to quantify the presence of outliers in the population. In particular, we find that the most massive detection from the first two observing runs is consistent with the $\sim 45\ M_\odot$ mass cutoff, and is a fairly typical member of the black hole population as opposed to a population outlier.

This chapter was written in collaboration with Will Farr and Daniel Holz, and published in the Astrophysical Journal Letters [107]. We later applied some of the outlier tests developed in this work to other potential outliers in the population, including the asymmetric binary black hole system GW190412 (see Chapter 5).

We thank Tom Callister for his helpful comments on the manuscript. MF was supported by the NSF Graduate Research Fellowship Program under grant DGE-1746045. MF and DEH were supported by NSF grant PHY-1708081. They were also supported by the Kavli

Institute for Cosmological Physics at the University of Chicago through NSF grant PHY-1125897 and an endowment from the Kavli Foundation. Part of this work was completed at the Kavli Institute for Theoretical Physics, supported by NSF grant PHY-1748958. DEH also gratefully acknowledges support from the Marion and Stuart Rice Award.

6.1 Abstract

Advanced LIGO and Virgo detected ten binary black holes (BBHs) in their first two observing runs (O1 and O2). Analysis of these events found evidence for a dearth of BBHs with component masses greater than $\sim 45\ M_\odot$, as would be expected from pair-instability supernovae. Meanwhile, a standalone analysis of the merger GW170729 found its primary mass $m_1 = 51.2^{+16.2}_{-11.0}\ M_\odot$, which appears to be in contradiction with the existence of a limit at $\sim 45\ M_\odot$. In this work, we argue that the masses of individual events can only be evaluated with reference to the full population. When GW170729 is analyzed jointly with the remaining detections, its inferred primary mass tightens considerably, to $m_1 = 38.9^{+7.3}_{-4.5}\ M_\odot$. In the presence of noise, apparent outliers in the detected distribution are inevitable. We discuss methods of distinguishing between statistical fluctuations and true population outliers using posterior predictive tests. Applying these tests to O1 and O2, we find that the ten detections are consistent with even the simplest power-law plus maximum-mass model considered by the LVC, supporting the claim that GW170729 is not a population outlier. We also provide non-parametric constraints on the rate of high-mass mergers and conservatively bound the rate of mergers with $m_1 > 45\ M_\odot$ at $2.8^{+5.4}_{-2.0}\%$ of the total merger rate. After 100 detections like those of O1 and O2 from a population with a maximum primary mass of $45\,M_\odot$, it would be common for the most massive system to have an observed maximum-likelihood mass $m_1 \gtrsim 70\,M_\odot$.

6.2 Introduction

A major goal of gravitational-wave (GW) astronomy is to learn about the formation and evolutionary mechanisms of binary black hole (BBH) mergers, such as those detected by Advanced LIGO [1] and Virgo [23]. There are many proposed formation channels for BBHs, including isolated evolution [91, 39, 253, 94, 45, 238, 165, 233], dynamical formation [183, 139, 212, 32, 65, 210, 218, 257, 83], and primordial origin [48, 117], and if several formation channels are active at once, the population of merging BBHs may consist of distinct sub-populations. These sub-populations may differ in their shape of the mass distribution and spin distribution, as well as the merger rate (and its evolution with redshift).

Previous studies have explored methods of distinguishing between different formation channels using GW observations of BBHs, including fitting for the mixture fraction (or branching ratio) between various sub-populations [237, 214, 256, 252, 236, 50]. One proposed sub-population includes second-generation mergers, which occur when at least one of the component BHs in a binary is itself the product of a previous merger. Second-generation BHs are expected to have a characteristic distribution of dimensionless spin magnitudes that peaks at $a \sim 0.7$ and a mass distribution that extends into the "pair-instability" mass gap starting at $\sim$ 40–70 $M_\odot$ [106, 120, 210]. On the other hand, "first-generation" BHs that form directly from stellar collapse are expected to avoid the mass range between $\sim$ 50–120 $M_\odot$. When a stellar core lies in this mass range, it is predicted to explode as a (pulsational) pair-instability supernova, and either leave behind no remnant, or lose mass sufficiently so that the BH remnant lies below the gap [115, 209, 127]. The spins of BHs formed from stellar collapse are less certain [191, 40, 116].

Second-generation mergers are possible (and generally expected to occur) exclusively in dense stellar environments such as globular clusters; therefore, the existence of this population is an important discriminator between dynamical and isolated formation. An additional proposed sub-population consists of gravitationally-lensed GW signals, for which the lensing

magnification causes a bias in the inferred luminosity distance and the unredshifted, source-frame masses if not properly accounted for [58]. Therefore, gravitationally-lensed events, even if they originate from the same formation channel as the unlensed events, would appear as a sub-population of erroneously high-mass, low-distance events [76, 194, 197, 51, 125].

The LVC detected ten BBHs in its first two observing runs [17]. With these ten detections, [19] fit simple parameterized models to the mass, spin, and redshift distribution of the BBH population, assuming that all detections belong to the same population[1]. The assumption of a single population was justified by a leave-one-out analysis, which shows explicitly that excluding GW170729, the most "unusual" event (in terms of having the highest mass, spin, and distance), from the analysis does not significantly impact the inferred mass, spin, and redshift distributions at a level beyond the statistical uncertainties. [149] and [66] also found that there is insufficient evidence to claim GW170729 as a population outlier by specifically comparing the hypothesis that it belongs to a population of second-generation, as opposed to first-generation, mergers based on the expected mass and spin distributions under the two scenarios.

In this paper we examine in more detail whether the assumed single-population mass distribution is a good fit to the data from the first ten detections, with a focus on the inferred "maximum mass," or lower edge of the pair-instability mass gap. We argue that there are no convincing outliers among the O1 and O2 detections, implying that a single-component population model is sufficient to fit the data. As the number of detections increases, we expect standard statistical fluctuations to produce individual events with significant posterior support inside the so-called upper mass gap; GW170729 may be an example of such a fluctuation. Additionally, we forecast the masses of future detections based on the population of BBHs from O1 and O2, and explore the masses that would be required to identify a BBH detection as a true population outlier. Such an outlier may belong to an alternate

1. We note that in addition to the LVC-published detections of [17], new BBH detections in the O1 and O2 data have been reported by [248], [247] and [196]

population consisting of, e.g., second-generation mergers, or otherwise indicate that the assumed population model provides an insufficient description of the data. The methods discussed here can be extended to classify any outliers, including in the spin, mass-ratio, or redshift distribution.

The parametric models discussed above and used in [19] are designed to incorporate a high-mass feature inspired by the predicted pair-instability mass-gap, whether it is a sharp cut-off to the power-law [Models A and B of 19] or a Gaussian component (Model C). To constrain the rate of high-mass BBH mergers in a model-agnostic way, we apply the non-parametric method of [178] to the ten BBHs from O1 and O2. This method models the mass distribution as a binned histogram in the m_1–m_2 plane, with a smoothing prior on the bin heights. In contrast to the parametric models, the non-parametric fit a priori prefers smoothness over sharp features, providing a conservative upper limit on the rate of high-mass mergers.

The remainder of the paper is organized as follows. In Section 6.3, we demonstrate how a population analysis provides updated inference on the event-level parameters, yielding a tighter measurement on the masses of GW170729 in particular. In Section 6.4, we use posterior predictive distributions to evaluate the goodness-of-fit of a model to observations, focusing on high-mass outliers. In Section 6.5, we apply the non-parametric histogram model of [178] to the ten LVC events and compare the inferred mass distribution to the parametric inference. We conclude in Section 6.6. Additional analysis details are provided in the Appendix.

6.3 Population Prior on Individual Events

A population analysis is concerned with fitting features that are common across the population members; in this case, BBH merger events. We assume that the event-level parameters $\{\theta_i\}$ (e.g. the component masses of the BBH event i) follow a probability distribution func-

tion $p_{\rm pop}\left(\theta_i \mid \Lambda\right)$, where Λ are the population-level hyperparameters (e.g. the power-law slope of the primary mass distribution). A hierarchical Bayesian analysis simultaneously fits for the event-level parameters θ_i and the population-level hyperparameters Λ from the data [134, 176, 180]. In the presence of selection effects and measurement uncertainty, the joint posterior probability distribution of the event-level parameters $\{\theta_i\}$ and the population parameters Λ given the data $\{d_i\}$ from N independent events is given by [169, 179]:

$$p(\{\theta_i\}, \Lambda \mid \{d_i\}) \propto \pi(\Lambda) \prod_{i=1}^{N} \frac{\mathcal{L}(d_i \mid \theta_i) p_{\rm pop}(\theta_i \mid \Lambda)}{\int P_{\rm det}(\theta) p_{\rm pop}(\theta \mid \Lambda) {\rm d}\theta}, \tag{6.1}$$

where $\pi(\Lambda)$ is the prior on the population hyperparameters, $\mathcal{L}(d_i \mid \theta_i)$ is the event-level likelihood, and $P_{\rm det}(\theta)$ is the probability of detecting a piece of data from a merger with true parameters θ, as discussed in detail below.

In a population analysis, we usually marginalize Eq. 6.1 over the event-level parameters θ_i to recover the posterior of the population parameters Λ. Meanwhile, when analyzing data from an individual event ("parameter estimation," or PE), the posterior on the event-level parameters,

$$p(\theta_i \mid d_i) \propto \mathcal{L}(d_i \mid \theta_i) \pi(\theta_i), \tag{6.2}$$

is typically calculated using a default uninformative prior $\pi(\theta_i)$, rather than a population prior $p_{\rm pop}(\theta_i \mid \Lambda)$. The mass estimates in [17], for instance, are obtained using priors that are flat in detector-frame component masses. Alternatively, one can marginalize Eq. 6.1 over the population-level parameters Λ, and get a new posterior on the event-level parameters for each detection. Therefore, by calculating a joint posterior on the population-level and event-level parameters simultaneously, a hierarchical Bayesian analysis in effect replaces the default uninformative priors from PE with a population prior, yielding an informed posterior on the event-level parameters.

Since the first BBH detections, a variety of hierarchical analyses have been carried out to

fit the BBH mass, spin, and redshift distributions to phenomenological parametric models [3, 159, 105, 240, 215, 109, 255]. In this work, we focus on the population analysis of [19]; specifically the mass distribution, which [19] fit according to three power-law based models: Models A, B, and C. Model A has two free parameters: the power-law slope for the primary mass distribution and the maximum mass, and fixes the minimum mass to 5 $M_\odot$ and the conditional distribution for the secondary mass to be flat between the minimum mass and m_1. Model B is a generalization of Model A with two additional parameters: the minimum mass and the power-law slope of the secondary mass (or equivalently, mass ratio) distribution as conditioned on the primary mass. Model C generalizes Model B by tapering the low- and high-mass ends of the mass function (as opposed to the sharp cutoffs of Models A and B) and allowing a high-mass Gaussian component on top of the power-law in primary mass. For each model, one can use the posterior samples for the population parameters Λ to recover the population-informed posteriors for individual event parameters θ_i according to Eq. 6.1.[2]

In certain cases, the population analysis returns a noticeably different posterior for single-event parameters compared to the posterior returned from PE. For example, for the most massive event of O1 and O2, GW170729, the population analysis implies a much tighter prior on the masses compared to the uninformative priors of PE. The population-informed posteriors on m_1 and q for GW170729 under Model B are shown in Figure 6.1.[3] The Model B population analysis implies that the primary mass is $m_1 = 38.9^{+7.3}_{-4.5}\ M_\odot$ and the mass ratio is $q = 0.95^{+0.05}_{-0.18}$ compared to $m_1 = 51.2^{+16.2}_{-11.0}\ M_\odot$ and $q = 0.63^{+0.32}_{-0.26}$ under the default uninformative priors. Even under Model C, which has a high-mass Gaussian component instead of the sharp cutoff of Model B, we find that the population-informed primary mass

2. The posterior samples on the population hyper-parameters from [19] are publicly available at https://dcc.ligo.org/LIGO-P1800324/public and the PE samples for individual events from [17] are available at https://dcc.ligo.org/LIGO-P1800370/public.

3. We use the parameter estimation results derived using the `IMRPhenomPv2` [148] waveform approximant throughout. PE posteriors and population results with the `SEOBNRv3` [202] approximant are available as well, and there are no significant differences between the two.

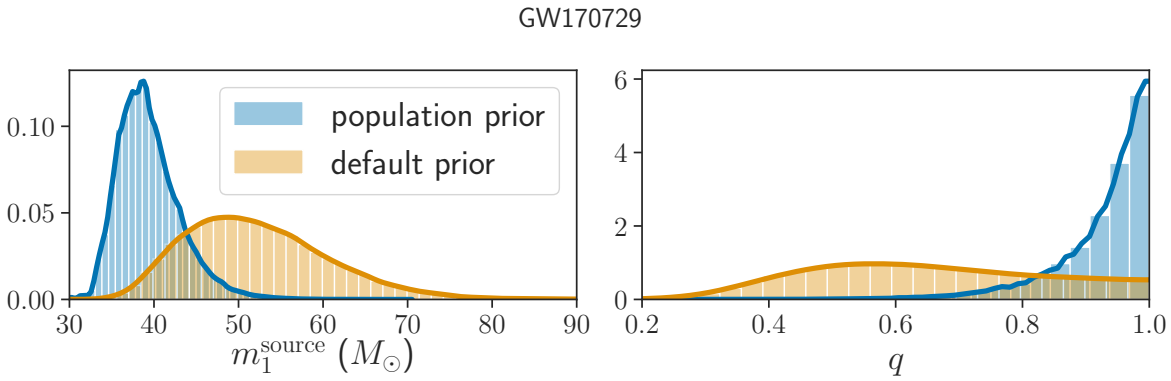

Figure 6.1: **Primary mass (left panel) and mass ratio (right panel) of the BBH event GW170729** under the default (flat in detector-frame masses) prior (orange) [although i think i would prefer green] versus the prior implied by the Model B population analysis of [19]. The population analysis strongly constrains the maximum mass of the population to $\lesssim 50\ M_\odot$, and favors near-unity mass ratios, which implies that, if we believe the parametric model is a reasonable description of the population, the primary mass and mass ratio of GW170729 are relatively well-constrained. Alternatively, the analysis with the default prior, which ignores the rest of the detected population, ascribes a high probability for an outlier value to the mass.

of GW170729 is $m_1 = 37.9^{+7.5}_{-6.4}\ M_\odot$. Under either model, the population analysis implies that the primary mass of GW170729 is less than 50 $M_\odot$ at 99% credibility, in contrast to the results with uninformative priors.

6.4 Evaluating tension between model and data

The above section assumes that the full set of observations can be adequately described by a given model, and it therefore makes sense to impose population priors and recalculate the posteriors on the parameters of each event. However, we often want to explicitly check whether the model fits the data sufficiently and whether the observations are consistent with one another under the model. In this section, we detail various methods of carrying out goodness-of-fit and outlier identification tests, and apply them to the BBH mass distribution

fits from [19].

6.4.1 Definitions and assumptions

In a hierarchical population analysis, there are three levels at which we can perform a model goodness-of-fit/consistency test. The highest level consists of the population parameters Λ and their inferred values from the data. If fitting the population model separately on different subsets of events yields posteriors on the population parameters that are in significant tension with one another, one can conclude that the model does not fit all events as a single population. An example of this test was carried out for Model A of [19] with a leave-one-out analysis, in which the population hyper-parameters were fit with and without GW170729. Comparing the posteriors on the hyper-parameters with and without GW170729 show that excluding GW170729 from the fit results in statistically consistent posteriors, leading to the conclusion that GW170729 is not a population outlier.

The second level of a hierarchical analysis consists of the (true) event-level parameters $\{\theta_i\}$. Following [19], we define the posterior population distribution as a probability density function (pdf) on the true parameters θ of an (unobserved, and potentially unobservable) system that belongs to the population, given the data we have already observed:

$$p(\theta \mid \{d_i\}) = \int p_{\rm pop}(\theta \mid \Lambda) p(\Lambda \mid \{d_i\}) \mathrm{d}\Lambda, \tag{6.3}$$

where the posterior $p(\Lambda \mid \{d_i\})$ is given by marginalizing Eq. 6.1 over the event-level parameters $\{\theta_i\}_{i=1}^{N}$ of the N detected events. The pdf on the true parameters θ of a *detection* must take into account selection effects by weighting each θ by the detectability of the data $\tilde{d}$ that it would produce in our detectors:

$$\begin{aligned} p(\theta, \mathrm{det} \mid \{d_i\}) &= p(\theta \mid \{d_i\}) P_{\rm det}(\theta) \\ &= p(\theta \mid \{d_i\}) \int P_{\rm det}(\tilde{d}) p(\tilde{d} \mid \theta) \mathrm{d}\tilde{d}. \end{aligned} \tag{6.4}$$

Note that the first term outside of the integral above is the posterior population distribution given by Eq. 6.3. The term $p(\tilde{d} \mid \theta)$ takes into account the measurement uncertainty in going from the true parameters of the system θ to the data $\tilde{d}$, and the term $P_{\rm det}(\tilde{d})$ accounts for the fact that only some pieces of data are detectable. Throughout, we assume that the detectability of a piece of data, $P_{\rm det}(\tilde{d})$, is deterministic, meaning it is always 0 or 1, depending on whether the data pass a (known) detection threshold; this is discussed in more detail in Appendix 6.7.1. The term $P_{\rm det}(\theta)$ also appears in Eq. 6.1. Sometimes $VT(\theta)$, the sensitive spacetime volume to a system with parameters θ, appears in place of $P_{\rm det}(\theta)$, as these terms are proportional to each other assuming the merger rate is constant in comoving volume and source-frame time. Given a collection of events with data $\{d_i\}$ and a population model, we can calculate the above pdf for the true parameters of detected events (Eq. 6.4). Comparing the true parameters $\{\theta_i\}$ of the detected events (as inferred jointly with the population hyperparameters via Eq. 6.1) against the posterior predictive pdf $p(\theta, \mathrm{det} \mid \{d_i\})$ provides another measurement of the consistency of the population model with the observations.

The third and final level of a hierarchical analysis consists of the data; this is the level on which we will evaluate the population fits for the remainder of this work. By folding in measurement uncertainty as well as the detection efficiency, we arrive at a probability distribution on the data $\tilde{d}$ from a future detection, rather than on the true parameters. We refer to this as the posterior predictive distribution:

$$p(\tilde{d}, \mathrm{det} \mid \{d_i\}) = \int P_{\rm det}(\tilde{d}) p(\tilde{d} \mid \theta) p(\theta \mid \{d_i\}) \mathrm{d}\theta. \tag{6.5}$$

A detected piece of data d from a BBH merger is a timeseries consisting of the GW signal $h(t)$, which, assuming GR, is fully described by the source parameters θ and detector noise $n(t)$. Therefore, we shall identify the data d with the set of observed (maximum-likelihood) source parameters $\theta_{\rm obs}$, by which we mean the true source parameters offset by some mea-

surement noise: $d = \{{m_1}^{\rm obs}, {m_2}^{\rm obs}, {z}^{\rm obs}, \ldots\} = \theta_{\rm obs}$. Once an event is detected, the PE analysis returns a posterior pdf on the source parameters $p(\theta \mid \tilde{d})$ [245]. In referring to the observed parameters $\theta_{\rm obs}$ in this work, we mean the maximum likelihood point-values as returned by PE.[4] The details of how we simulate mock GW datasets with realistic measurement uncertainty in order to compute $P_{\rm det}(\tilde{d})$ and $p(\tilde{d} \mid \theta)$ are found in Appendix 6.7.1.

We stress that *noise fluctuations will generally cause us to estimate erroneously large source-frame masses.* A notable consequence of the detection threshold is that near-threshold sources are detectable only if a favorable noise fluctuation pulls the observed SNR above the detection threshold. Because the distance measurement correlates with the observed SNR, the systematic shift to larger SNR caused by the detection threshold implies that the observed redshift (as inferred from the distance) tends to be systematically shifted towards smaller values. This in turn implies that the observed source-frame masses, which are inferred by un-redshifting the detector-frame masses, are preferentially shifted towards larger values.

Although extreme noise fluctuations are intrinsically rare, they are statistically guaranteed to affect some fraction of detections. The larger the sample of detections, the larger the fluctuation it contains.

6.4.2 Application to the O1+O2 population

Posterior Predictive Goodness-of-Fit

In this subsection, we calculate the posterior predictive mass distribution for the BBH population from O1 and O2 and explore whether the population models provide an adequate fit to the ten BBH observations. We focus on their primary masses, and more specifically, the primary mass of GW170729.

4. The maximum-likelihood values do not correspond to the peak of the posteriors shown for event parameters in [17] because a nonflat prior $\pi(m_1, m_2, z)$ is used. The default prior in the event-level analysis is proportional to the square of the luminosity distance, which places more weight at high z than a flat prior, so the maximum-likelihood source-frame mass values are larger than the maximum a posteriori values.

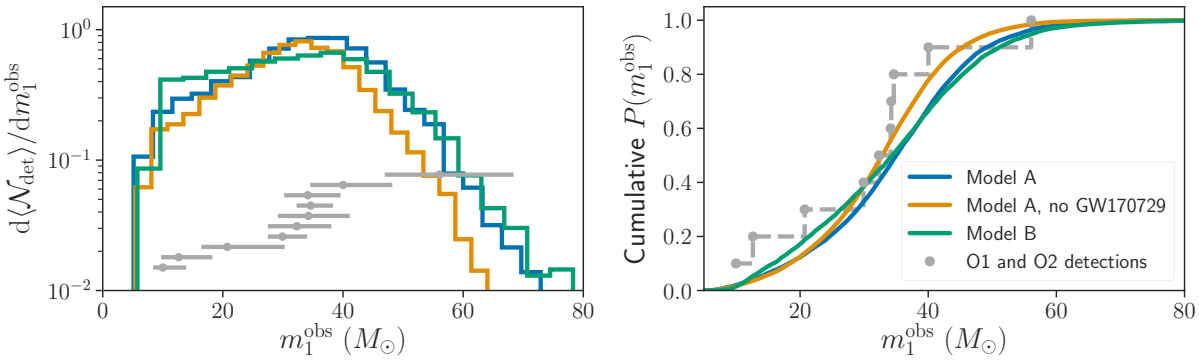

Figure 6.2: *Left panel:* **The posterior predictive distribution for the number of detections per *observed* primary mass bin during O1 and O2**, based on the fits to Model A and Model B using all ten BBHs (blue and green lines) and the nine BBHs excluding GW170729 (yellow). The gray error bars show the maximum-likelihood points and 90% credible intervals on the masses of each of the ten BBHs; the vertical placement of these error bars is arbitrary. The observed mass is defined as the maximum likelihood estimate of m_1, and is predicted according to the synthetic detection and PE process described in the text. *Right panel:* The cumulative posterior predictive distribution, or the probability that an observed mass is less than m_1^{obs}, inferred from the detections and the given population model, compared to the empirical distribution function from the ten detections in gray (with the gray points denoting the maximum likelihood m_1 estimates). The agreement between the observations and each model can be quantified by the distances between the gray points and the colored curve of interest, as calculated in the text. The predictions of the population model match the observations fairly well, and GW170729 does not appear to be an outlier even when excluding it from the calculation of the posterior predictive distribution (yellow curve).

The posterior predictive distribution (Eq. 6.5) for $m_1^{\rm obs}$, given the events from O1 and O2 and the assumed mass Models A and B from [19], is shown in Figure 6.2. We focus on Models A and B, the simple power-law models, because they have fewer free parameters than Model C, and we wish to check whether these few parameters sufficiently fit the data. We show the posterior predictive distributions as inferred from all ten BBHs, as well as under Model A excluding GW170729. The left panel in Figure 6.2 shows the expected number of detections per $m_1^{\rm obs}$ bin during O1 and O2 according to the model; this is based on the inferred merger rate together with the shape of the mass distribution.

We note that although Model A (B) predicts a sharp cutoff at $m_{\rm max} = 41.6^{+9.6}_{-4.3}\ M_\odot$ ($m_{\rm max} = 40.8^{+11.8}_{-4.4}\ M_\odot$), and all three models considered in [19] predict that 99% of BBHs have $m_1 < 45\ M_\odot$, Model A (B) predicts that 18% (20%) of detected systems have an *observed* primary mass $m_1^{\rm obs} > 45\ M_\odot$. This is because out of the underlying population of BBHs, more massive systems are more likely to be detected, and out of those that are detected, statistical fluctuations can push the observed (maximum likelihood) $m_1^{\rm obs}$ to values that are oftentimes significantly larger than the true m_1. Recall that among detected sources, statistical fluctuations are more likely to push the observed source-frame masses to larger values than smaller values, because sources near threshold are only detectable if a fluctuation increases their observed SNR, leading to a smaller inferred redshift and a larger inferred source-frame mass. The full likelihood distribution should still have nonzero support at the true value, but for very large statistical fluctuations, the support at the true value may be very small and difficult to resolve.

From Figure 6.2, one can visually compare the number and observed masses of the O1 and O2 detections (gray errorbars) to the model predictions; this serves as a posterior predictive check that the model fits the data sensibly. The gray points at the center of the errorbars denote the maximum likelihood $m_1^{\rm obs}$ for each event; these points are the values that, according to the model, should be representative draws from the posterior predictive

distributions. (The errorbars denote the 90% symmetric credible intervals from the full m_1 posteriors and are shown only for reference).

The right panel of Figure 6.2 shows the cumulative, normalized versions of the posterior predictive distributions in the right panel (the colored curves) compared to the empirical distribution function (edf; the gray points). The edf is a cumulative histogram of the maximum likelihood m_1, $\hat{m}_{1,i}^{\rm obs}$, for each event i, defined as:

$$\hat{F}_n(m_1^{\rm obs}) = \frac{1}{n}\sum_i^n I[\hat{m}_{1,i}^{\rm obs} < m_1^{\rm obs}], \tag{6.6}$$

where n is the number of events and I is the indicator function which evaluates to 1 if its argument is True and zero otherwise.

As seen in the right panel of Figure 6.2, the edf appears to follow the posterior predictive cdfs for Models A and B, with perhaps slightly more low-mass and high-mass detections than predicted under the simple power-law models. The relative lack of intermediate-mass detections (and comparable abundance of detections at the low- and high-mass ends of the $m_{\rm min}$-$m_{\rm max}$ range) can also be seen in the fit to Model C (the tapered power-law with a high-mass Gaussian) of [19]. Under Model C, the only model that allows for deviations from a pure power-law in m_1, [19] find that the data mildly prefers a merger rate that decreases at intermediate m_1 and then rises again at $m_1 \sim 30\ M_\odot$ (see the top panel of their Figure 2). [19] conclude that this preference for a power-law deviation is not statistically significant because all three models A, B, and C predict distributions of m_1 that overlap within the 90% statistical uncertainties. [19] also report Bayes' factors between all three models and find that a deviation from a power-law is preferred by a factor of $e^{1.92} \approx 7$, although as usual, the values of Bayes' factors are sensitive to the priors chosen for the model hyperparameters. Our posterior predictive checks for the pure power-law Models A and B do not rely on an explicit comparison to an alternate model, but our conclusions agree with [19]

that the observed m_1 distribution is consistent with these power-law models well within the 90% level.

More quantitatively, the distance between the edf and the cumulative probability distributions (cdf) predicted from the models as fit to the data is a measure of "goodness-of-fit," or how well each model explains the data. This is the basis of the well-known Kolmogorov-Smirnov (KS) [155, 227] and Anderson-Darling (AD) statistics [28].[5] Restricting ourselves only to the m_1 distribution, we compute the KS and AD statistics as follows. We first construct a null distribution by repeatedly drawing sets of ten observations from the posterior predictive distribution of interest and calculating the KS and AD statistic for each set. Comparing the KS and AD statistics for our set of ten observations against the statistics of the null distribution, we find that the KS statistic between the edf and the Model B posterior predictive cdf lies at 84% of the null distribution, while the AD statistic lies at 79% of its null distribution. For Model A, the KS (AD) statistic is at 87% (82%) of its null distribution. We also compute these statistics between the edf for all ten events and the Model A posterior predictive cdf as inferred without GW170729. This does not affect the statistics significantly, with the KS statistic corresponding to 69% of the null distribution and the AD statistic at 71% in this case, which indicates that the observed primary mass of GW170729 is consistent with the m_1 distribution as inferred without it.[6] *In summary, Models A and B are all found to be consistent with the observed population.*

Note that the edf is not uniquely defined in two or more dimensions, and Figure 6.2 shows

5. The KS statistic is the maximum distance between the cdf of the model and the edf derived from the data (or two edfs for two different datasets) while the AD statistic is a weighted average of the distance between each point in the sample and the cdf. The KS statistic is mainly sensitive to differences between the centers of the two distributions while the AD statistic is more sensitive to differences in the tails of the distribution [235].

6. Counterintuitively, the shift towards smaller KS and AD statistics seems to indicate that the m_1 posterior predictive distribution as inferred from only nine events actually provides a slightly better fit to the observed primary masses of all ten events; however, the shift in these statistics is not significant, and illustrates that this is a simplified one-dimensional test that is only meant to check consistency between data and model and not to properly fit a model.

only how the model fits the observed m_1 distribution, ignoring all other parameters of BBH systems including the mass ratio, spin, and redshift. However, the edf-based test described above can be extended (non-uniquely) to an arbitrary number of dimensions as long as the algorithm for ordering the observations is fixed in advanced. The null distribution can still be calculated according to this fixed ordering algorithm, and, if desired, a p-value can be calculated to evaluate goodness-of-fit.

Quantifying Outliers

We now turn to the more specific issue of quantifying whether a particular observation is an outlier with respect to other events that make up the population. For example, in order to investigate the degree of tension between a particularly massive observation, such as GW170729, and the maximum mass inferred from the population analysis, it is useful to consider the posterior predictive distribution of the *maximum* observed primary mass out of N detections, $\max\left(\{m^j_{1,\mathrm{obs}}\}_{j=1}^{N}\right)$, based on the data from the remaining $N-1$ detections, $\{d_i\}_{i=1}^{N-1}$:

$$\begin{aligned} & p\left(\max\left(\{m^j_{1,\mathrm{obs}}\}_{j=1}^{N}\right)\middle|\,\{d_i\}_{i=1}^{N-1}\right) \\ =&\frac{\mathrm{d}}{\mathrm{d}m}\left(\left[\int_0^m p(m_{1,\mathrm{obs}}\left|\{d_i\}_{i=1}^{N-1}\right.)\mathrm{d}m_{1,\mathrm{obs}}\right]^{N-1}\right)\Bigg|_{m=\max\left(\{m^j_{1,\mathrm{obs}}\}_{j=1}^{N}\right)}, \end{aligned} \tag{6.7}$$

where the pdf in the integral is simply the posterior predictive distribution. Equation 6.7 follows from the fact that given N random draws X_i from a fixed pdf $p(x)$ with corresponding cdf $P(X_i < x)$, the maximum draw Y will follow a cdf:

$$P(Y<x) = P(X_i<x)_{\forall i\in[1,N]} = P(X_i<x)^N. \tag{6.8}$$

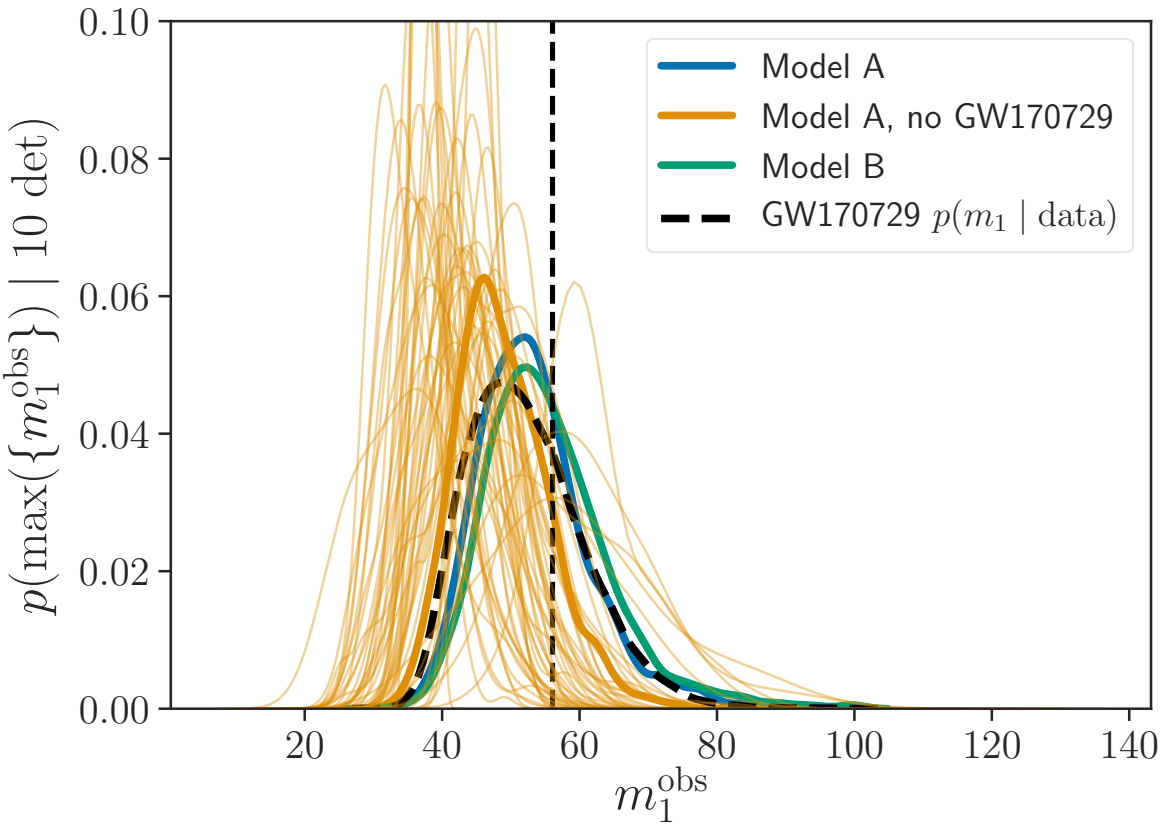

Figure 6.3: **Posterior predictive distribution of the maximum *observed* mass out of ten detections** as inferred from the detections and the population model of interest (bold, solid colored curves). The observed mass refers to the maximum likelihood m_1 value of a detected event as predicted according to a synthetic detection and PE process (see text). The thin orange curves show mock PE posteriors for 50 random events drawn from the bold orange curve, representing 50 examples of posteriors for the most massive m_1 that we expect to detect based on the fit to Model A from the nine detections excluding GW170729. For comparison, the posterior for the primary mass of GW170729 is shown (dashed black curve) with the maximum-likelihood value (vertical dashed line). Visually, the GW170729 m_1 posterior appears consistent with the thin orange curves. Quantitatively, comparing its maximum-likelihood value to the bold orange curve shows that the primary mass of GW170729 is consistent with the population as inferred from the other nine events at the 86% level.

The posterior predictive distributions for the maximum observed mass out of ten detections is shown in Figure 6.3. The maximum likelihood primary mass $m_1^{\rm obs}$ for GW170729 (vertical dashed line) is consistent at the 86% level with the posterior predictive distribution on the maximum of ten primary mass observations, as inferred from Model A and the remaining nine observations (orange solid curve). The light orange curves show mock posteriors on m_1 for fifty events drawn from the solid orange curve, and we can see that the m_1 posterior for GW170729 (dashed black curve) looks like a typical observation. When using GW170729 itself in calculating this posterior predictive distribution, its primary mass is consistent with being the maximum of ten observations at 69% under Model A (blue curve) and 59% under Model B (green curve). Based on this analysis, we conclude that there is no evidence for tension between the primary mass of GW170729 and the remaining nine observations under Models A and B, in agreement with the leave-one-out analysis of [19].

As the sample of BBH detections grow, we expect to see more extreme statistical fluctuations, so that we observe primary masses $m_1^{\rm obs}$ that are much higher than the true maximum BH mass. In Figure 6.4, we show, as a function of the number of detections, the most massive $m_1^{\rm obs}$ that we expect in the sample, based on the fit to Model B with the ten O1/O2 detections. We find that our predictions for the maximum observed mass after N detections matches expectations from analytic Gaussian scaling. Assuming Gaussian observational uncertainties of width σ, with N detections, some fraction fN of systems will have masses within 1-σ of the true maximum mass, where f depends on the shape of the mass distribution convolved with selection effects. The largest observed mass will following a scaling $\max m_1^{\rm obs} \sim \max m_1 + \sigma\sqrt{2\log(fN)}$; this scaling is observed empirically in Figure 4. In other words, the most extreme noise fluctuation in a sample of size N is typically observed at $\sqrt{2\log N}$ standard deviations above its true value. *It would be common that by the time we have 100 detections, even if they are all described by Model B with a sharp cutoff at* $m_{\rm max} = 40.8^{+11.8}_{-4.4}\ M_\odot$*, that we will observe a system with an apparent mass of*

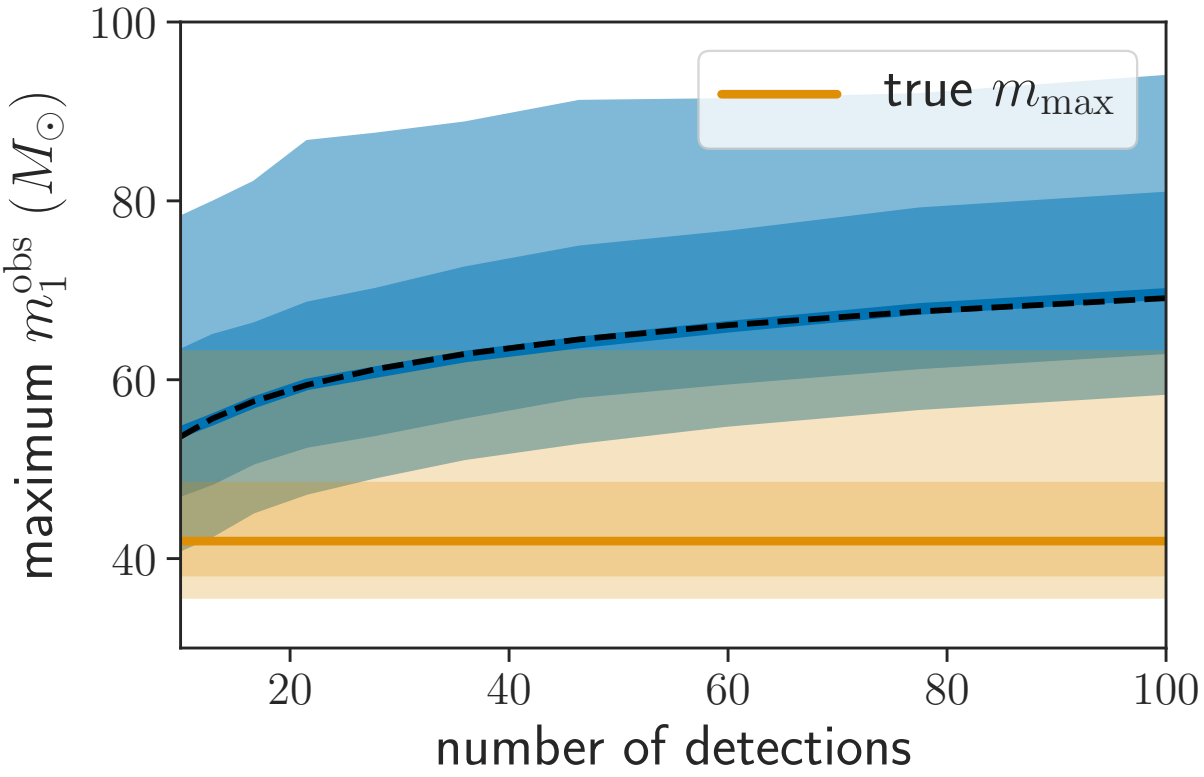

Figure 6.4: **Maximum mass we expect to observe** as a function of number of detections (blue), as inferred from the Model B fit to the ten O1/O2 detections (which predicts the true $m_{\rm max} = 40.8^{+11.8}_{-4.4}\ M_\odot$, shown in orange). The observed mass $m_1^{\rm obs}$ represents the maximum likelihood value that would be inferred for the detected system, taking into account measurement uncertainty and selection effects. The solid line denotes the median and dark and light bands denote the 68% and 95% credible intervals of the posterior predictive distribution. As the number of detections increases, the largest noise fluctuation in the sample will become more extreme. Furthermore, because of the SNR threshold, these noise fluctuations statistically lead to larger inferred masses. The blue curves are well fit by $\max m_1^{\rm obs} \simeq \max m_1 + \sigma\sqrt{2\log(fN)}$ with $\sigma \sim 17\ M_\odot$ (see text); this fit is shown by the black dashed line.

$m_1^{\rm obs} \sim 70\,M_\odot$.

6.5 Non-parametric Constraints on the Rate of Mass-Gap Mergers

To explore the rate of high-mass mergers in a less parametric way, we follow the binned-histogram analyses of [112] and [178]. We model the rate of BBH mergers on the two-dimensional m_1–m_2 plane, $\mathcal{R}(m_1, m_2) = \frac{dN}{dm_1 dm_2 dV_c dt}$ as piecewise constant in 9×9 logarithmically-spaced mass bins between 3 and 150 $M_\odot$. The height of each mass bin $\mathcal{R}_{ij}$ represents the merger rate in that bin. We take the prior on the bin heights to be a squared-exponential Gaussian process (GP) where the relative means μ_{ij} of the bin heights follow a fixed shape $p(m_{1,i}, m_{2,j})$ and the length scales in $\log(m_1)$ and $\log(m_2)$ are fit from the data (see Appendix 6.7.2 for more details). We consider two different shapes for the mean merger rate per bin in the GP prior: a "power-law" shape prior and "flat-in-log" shape prior. These are motivated by the two fixed-parameter models that the LVC used to calculate BBH merger rates in O1 and O2 [3, 17]. In the power-law shape prior, we have:

$$p(m_1, m_2) \propto \frac{m_1^{-2.35}}{m_1 - M_{\rm min}}, \tag{6.9}$$

where $m_{\rm min} = 3\ M_\odot$. The flat-in-log shape prior is simply:

$$p(m_1, m_2) \propto \frac{1}{m_1 m_2} \tag{6.10}$$

Figure 6.5 shows the merger rate $\mathcal{R}(m_1)$ as inferred under the two different priors, where:

$$\mathcal{R}(m_1) = \int_3^{m_1} \mathcal{R}(m_1, m_2) dm_2. \tag{6.11}$$

For comparison we also show the results from the parametrized Model B fit. We note that

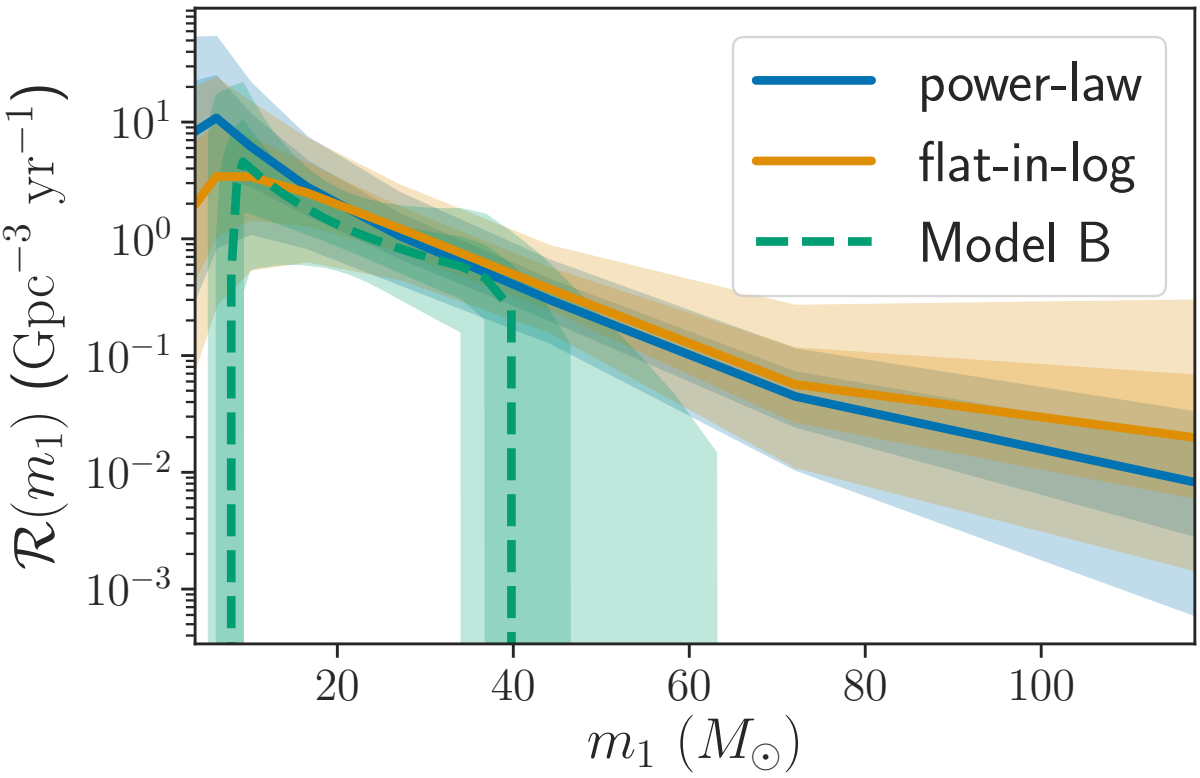

Figure 6.5: **The non-parametric constraints on the merger rate distribution** for m_1 from the binned histogram model under the two shape priors (flat-in-log prior in blue and power-law prior in orange). The solid lines denote the median rate and the dark and light shaded bands denote 68% and 95% symmetric credible intervals. In green we show the merger rate as a function of m_1 for the parametric mass distribution Model B; this model includes a maximum mass cutoff as one of the parameters. In the low-mass region $m_1 \lesssim 45\ M_\odot$, the non-parametric model under both priors agrees with the parametric model. Because of the lack of detections at high masses, the parametric Model B infers a tight constraint on $m_{\rm max}$ and the merger rate falls to zero, while the nonparametric model attempts to extrapolate smoothly to high masses under strong influence of the prior.

in the range $m_1 \lesssim 45\ M_\odot$, the inferred merger rates $\mathcal{R}(m_1)$ agree between all three models: the two shape priors in the nonparametric model as well as the parametric power-law model. (The difference for $m_1 < 5\ M_\odot$ is due to the prior on $m_{\rm min} > 5\ M_\odot$ for Model B, while the lowest mass bin in the nonparametric model starts at $3\ M_\odot$.) Beyond $\sim 45\ M_\odot$, the merger rate inferred under Model B drops sharply due to the $m_{\rm max}$ parameter, whose inferred value closely follows the mass of the most massive observed system [105]. The binned model, on the other, does not have a $m_{\rm max}$ parameter that lets the rate fall to zero, and instead has a prior that strongly favors smooth variations of the merger rate from mass bin to mass bin. In the mass bins with $m_1 \gtrsim 45\ M_\odot$, where there are no detections, the posterior on the merger rate smoothly transitions to following the prior on the bin heights. This is visible in Figure 6.5 as the merger rate $\mathcal{R}(m_1)$ inferred under the two different priors (blue and orange bands) start to diverge from one another at $m_1 \gtrsim 45\ M_\odot$, and the uncertainty for each one grows as well. This is a consequence of the GP smoothing prior. After enough detections, if the absence of high-mass events continues, the likelihood will overcome the smoothing prior and the posterior will reveal a sharp drop-off in the merger rate in the binned analysis, independently of the prior on the bin heights. This was demonstrated by [178] with simulated data in the context of the putative low mass gap between the binary neutron star (BNS) and BBH population. With only ten detections, the binned model provides a conservative upper limit on the rate of mergers with $45\ M_\odot < m_1 < 150\ M_\odot$ under the prior that the merger rate should not vary sharply between neighboring mass bins.

Under the flat-in-log prior, we infer the merger rate in the mass range $45\ M_\odot < m_1 < 150\ M_\odot$ to be $3.02^{+12.97}_{-2.28}$ Gpc^{-3} yr^{-1} (90% equal-tailed credible interval), or $1.79^{+2.30}_{-1.23}$ Gpc^{-3} yr^{-1} under the power-law prior. We infer the total merger rate over the $3\ M_\odot < m_1 < 150\ M_\odot$ range to be $42.50^{+68.12}_{-24.89}$ Gpc^{-3} yr^{-1} with the flat-in-log shape prior or $65.58^{+102.34}_{-41.82}$ Gpc^{-3} yr^{-1} with the power-law shape prior, implying that the rate of mergers with $m_1 > 45 M_\odot$ makes up $7.6\%^{+23.8}_{-6.0}$ (flat-in-log prior) or $2.8\%^{+5.4}_{-2.0}$ (power-law prior) of

the total merger rate. This is to be compared to the parametric models of [19], which all predict that less than 1% of mergers have $m_1 > 45\ M_\odot$. Unlike the parametric models with a maximum mass parameter, the binned-histogram model does not allow the high-mass merger rate to drop all the way to zero. The advantage of the nonparametric constraints is that, if there were a secondary population of BBH mergers that does not respect the maximum mass feature of the parametrized mass models (consisting, for example, of second-generation mergers that occupy the pair-instability mass gap), we still expect their merger rate to respect these nonparametric limits.

6.6 Conclusion

Focusing on the BBH mass distribution as inferred by [19], we have explored how individual events fit into a population analysis, especially in the presence of measurement uncertainty and selection effects. We have presented simple posterior predictive/goodness-of-fit checks to show consistency between the O1/O2 events and the power-law mass distribution models of [19]. In particular, GW170729, the most massive event of O1/O2, is not an outlier with respect to even the simplest power-law model with a sharp high-mass cutoff. When folding in the full information about the population, the primary mass of GW170729 is inferred to be $m_1 = 38.9^{+7.3}_{-4.5}\ M_\odot$ as compared to the inferred value under the uninformative PE priors, $m_1 = 51.2^{+16.2}_{-11.0}\ M_\odot$. We argue that as the number of detections increases, the sample will contain statistical fluctuations even more extreme than the one that likely gave rise to the observed mass of GW170729. These statistical fluctuations will lead us to preferentially overestimate the masses of individual events, unless we infer the individual events' masses jointly with the full population. In doing so, it will also be necessary to verify that the chosen population model fits the data sufficiently well. Although the simple parametrized models provide adequate fits to the BBH detections so far, we have presented nonparametric fits to the mass distribution based on the GP-regularized binned-histogram

model of [178]. Under this model, we place conservative upper limits on the rate of mergers with 45 $M_\odot < m_1 < 150\ M_\odot$, and find that these high-mass mergers make up at most $7.6\%^{+23.8}_{-6.0}$ of the total merger rate in the range $3\ M_\odot < m_1 < 150\ M_\odot$ for a flat-in-log shape prior on the mass distribution, or $2.8\%^{+5.4}_{-2.0}$ of the total merger rate for a power-law shape prior.

6.7 Additional Analysis Details

6.7.1 Mock observations

This section explains in greater detail how we calculate selection effects and simulate measurement uncertainty for mock observations. For a BBH with true parameters θ, we follow the simple prescription of [109] and [98] to assign realistic measurement uncertainty and compute $\theta_{\rm obs}$. Given the BBH's source-frame masses, spins, and redshift together with a PSD describing the noise of the detector, we can calculate the optimal SNR $\rho_{\rm opt}$ of the source, which is the SNR that it would have if it were optimally oriented face-on and directly overhead of the detector [71]. We assume a fixed cosmology described by the best-fit Planck 2015 parameters [24] to interchange between redshift and luminosity distance [33]. We fix the PSD to the "aLIGO Early High Sensitivity" noise curve from [18], which is representative for O1 and O2. We also fix BBH spins to zero in this calculation, since they have a negligible effect on the SNR calculation for population studies [19]. An isotropic distribution of sky positions and inclinations relative to a detector yields a distribution of true SNRs ρ in that detector; this distribution can be summarized by the angular projection term $0 \leq \Theta = \frac{\rho}{\rho_{\rm opt}} \leq 1$. The angular term Θ has a known distribution [103]. Therefore, for a BBH with intrinsic parameters $\{m_1, m_2, z\}$, we assign a single extrinsic parameter Θ drawn from this distribution. These four parameters together correspond to a true SNR ρ for the source.

Given $\{m_1, m_2, z, \Theta, \rho\}$ for a source, we assign measurement uncertainty as follows. We expect that in stationary, Gaussian noise, the matched-filter SNR ρ has unit variance [26]. We first draw an observed SNR $\rho_{\rm obs}$ from a normal distribution centered at the true SNR ρ:

$$\rho_{\rm obs} = \rho + N(0,1) \tag{6.12}$$

where $N(\mu, \sigma)$ is a random number drawn from a normal distribution with mean μ and standard deviation σ. We assume that sources are only detected if they pass an SNR threshold (in a single detector) $\rho_{\rm obs} > 8$; this would be identical to the semi-analytic selection effects calculation of [8, 3, 19] were it not for the inclusion of the noise term $N(0,1)$. To best approximate the mass measurement, we work with the detector-frame (redshifted) chirp mass:

$$\mathcal{M}_z = (1+z)\frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}}, \tag{6.13}$$

and symmetric mass ratio:

$$\eta = \frac{m_1 m_2}{(m_1+m_2)^2}. \tag{6.14}$$

The detector-frame chirp mass drives the leading-order GW frequency evolution during the inspiral and is thus the best-measured mass parameter for stellar-mass compact binary sources. We assume that the uncertainties $\{\sigma_{\mathcal{M}}, \sigma_\eta, \sigma_\Theta\}$ on the measured parameters scale inversely with $\rho_{\rm obs}$, so that:

$$\log \mathcal{M}_z^{\rm obs} = \log\left[\mathcal{M}_z + N(0, \sigma_{\mathcal{M}}/\rho_{\rm obs})\right], \tag{6.15}$$

$$\eta_{\rm obs} = N^{T[0,0.25]}(\eta, \sigma_\eta/\rho_{\rm obs}), \tag{6.16}$$

$$\Theta_{\rm obs} = N^{T[0,1]}(\Theta, \sigma_\Theta/\rho_{\rm obs}). \tag{6.17}$$

In the above expressions, $N^{T[a,b]}(\mu,\sigma)$ denotes a random number drawn from the truncated normal distribution. From $\mathcal{M}_z^{\rm obs}$ and $\eta_{\rm obs}$, we calculate the detector-frame component

masses $m_{1,z}^{\rm obs}$ and $m_{2,z}^{\rm obs}$:

$$m_{1/2,z}^{\rm obs} = \frac{M \pm \sqrt{M^2 - 4\eta M^2}}{2} \tag{6.18}$$

for $M = \mathcal{M}_z^{\rm obs}/\eta_{\rm obs}^{3/5}$. The observed redshift is inferred directly from the remaining parameters, via the observed luminosity distance $d_{\rm obs}$:

$$d_{\rm obs} = \frac{d_{\rm fid}\rho_{\rm opt}((1+z)m_1^{\rm obs}, (1+z)m_2^{\rm obs}, d_{\rm fid})\Theta_{\rm obs}}{\rho_{\rm obs}}, \tag{6.19}$$

where $d_{\rm fid}$ is an arbitrary fiducial luminosity distance (for given detector-frame masses, the SNR of a source scales inversely with its luminosity distance). The observed redshift $z_{\rm obs}$ is derived from $d_{\rm obs}$ by the cosmological redshift-luminosity distance relation, and once this is known, we infer the source-frame masses:

$$m_{1/2}^{\rm obs} = \frac{m_{1/2,z}^{\rm obs}}{1 + z_{\rm obs}}. \tag{6.20}$$

The observed values $m_1^{\rm obs}, m_2^{\rm obs}, z_{\rm obs}$ denote the maximum likelihood values of the parameters as extracted from the GW signal. To simulate full posterior distributions on these parameters, we use Eq. 6.12-6.19 as the likelihood for the observed parameters given true values ρ, $\mathcal{M}_z$, η, and Θ. We take samples on ρ, $\mathcal{M}_z$, η, and Θ from this likelihood and convert the samples to the space $m_{1,z}, m_{2,z}, d_L, \Theta$, on which we wish to set a prior:

$$\pi_{\rm PE}(m_{1,z}, m_{2,z}, d_L, \Theta) \propto p(\Theta) d_L^2. \tag{6.21}$$

This matches the default PE prior used by LIGO/Virgo in the O1 and O2 event analysis [17], where $p(\Theta)$ is the true distribution from which the Θ values are drawn, representing an isotropic distribution on the sky and source inclination. The change of variables from $\rho, \mathcal{M}_z, \eta, \Theta$ to $m_{1,z}, m_{2,z}, d_L, \Theta$ means we also have to divide out by the induced prior, given

by the Jacobian:

$$\left|\frac{\mathrm{d}\rho}{\mathrm{d}d_L}\frac{\mathrm{d}(\mathcal{M}_z,\eta)}{\mathrm{d}(m_{1,z},m_{2,z})}\right| = \frac{\Theta\rho_{\mathrm{opt}}(m_{1,z},m_{2,z},d_{\mathrm{fid}})d_{\mathrm{fid}}}{d_L^2}\frac{(m_{1,z}-m_{2,z})\eta(m_{1,z},m_{2,z})^{3/5}}{(m_{1,z}+m_{2,z})^2} \quad (6.22)$$

Once we reweight the $m_{1,z}, m_{2,z}, d_L, \Theta$ samples by Eq. 6.21 divided by Eq. 6.22, we can get posterior samples for the source-frame parameters by converting to m_1, m_2, z, Θ space.

We tune the σ parameters above to match the measurement uncertainties on masses and redshifts found by [251] when simulating full PE on injected signals. [251] found that for BBHs detected by Advanced LIGO/ Virgo at design sensitivity, the relative uncertainty (at the 90% credible interval) on the detector-frame masses is typically $\sim$ 40% and the relative uncertainty on redshift is typically $\sim$ 50%. However, for the majority of O1 and O2, only the two (co-aligned) LIGO detectors were operational, implying a reduced ability for the network to constrain the polarization of a source and break the distance-inclination degeneracy, and worsened redshift constraints. We find that for the O1 and O2 events, a more typical relative redshift uncertainty is 70% (for a 90% credible interval relative to the median value). We find that choosing $\sigma_{\mathcal{M}} = 0.08\rho_{\mathrm{thresh}}$, $\sigma_\eta = 0.022\rho_{\mathrm{thresh}}$ and $\sigma_\Theta = 0.21\rho_{\mathrm{thresh}}$ yields measurement uncertainties that match the widths of the O1 and O2 credible intervals and the expectations from [251]. We use $\rho_{\mathrm{thresh}} = 8$ throughout, as discussed above. The measurement uncertainty on Θ controls the measurement uncertainty on z according to Eq. 6.19. For simulating events for a 3-detector network at design sensitivity, we would use $\sigma_\Theta = 0.15$ to reflect the improved distance constraints of a 3-detector network with relative uncertainties of 50% rather than 70%. We note that the measurement uncertainty on the source-frame masses is a combination of the detector-frame mass uncertainty and the absolute redshift uncertainty, which is largest for sources at high redshift. Therefore, our predictions for distributions of m_1^{obs} based on distributions of m_1 are sensitive to assumptions regarding the underlying redshift distribution and the network sensitivity, which together determine the detected redshifts of the sources. The uncertainty in these assumptions is

subdominant to uncertainties in the population model.

6.7.2 Binned histogram likelihood

In this section, we provide additional analysis details regarding the binned-histogram fit to the mass distribution (Section 6.5). The total posterior for the rate $\mathcal{R}_{ij}$ in each mass bin, the parameters Γ governing the Gaussian process prior on the bin heights, and the true masses of the detected events is given by the inhomogeneous Poisson process likelihood:

$$p\left(\mathcal{R}_{ij}, \Gamma, \{m_1, m_2\} \mid d\right) \propto \left[\prod_{k=1}^{N_{\rm events}} p\left(d^{(k)} \mid m_1^{(k)}, m_2^{(k)}\right) \mathcal{R}(m_1^{(i)}, m_2^{(i)})\right] \times \exp\left[-\sum_{ij} \lambda_{ij}\right] p\left(\mathcal{R}_{ij}|\Gamma\right) p\left(\Gamma\right), \tag{6.23}$$

where $p\left(d^{(k)} \mid m_1^{(k)}, m_2^{(k)}\right)$ denotes the likelihood of the data for event k given component masses, marginalized over all other parameters of the system, $p\left(\mathcal{R}_{ij}|\Gamma\right)$ represents the GP prior on the bin heights and λ_{ij} denotes the expected number of detections in bin ij, folding in selection effects.

The GP prior on the (log) of the bin heights takes the form:

$$p(\log \mathcal{R}_{ij} \mid \Gamma) = N(\log \mathcal{R}_{ij} \mid \log(\mu + \mu_{ij}), \Sigma). \tag{6.24}$$

where $N(x \mid \hat{\mu}, \hat{\Sigma})$ denotes the multivariate normal probability distribution function on x with mean $\hat{\mu}$ and covariance $\hat{\Sigma}$. Meanwhile, μ is an overall scale factor and $\mu_{ij} = \log p(\log m_{1,i}, \log m_{2,j})$ is fixed to either the power-law or flat-in-log shape discussed in Section 6.5. Note that the point $(\log m_{1,i}, \log m_{2,j})$ denotes the center of the ijth bin. Finally, Σ is a covariance matrix. We use a squared exponential kernel for the covariance matrix,

and parameterize Σ as:

$$\Sigma_{ijkl} = (1-f)\sigma^2 \exp\left[-\frac{1}{2\tau^2}\Delta^2 \log m_{ijkl}\right] + f\sigma^2\delta_{ijkl}. \tag{6.25}$$

For numerical stability, the covariance Σ includes some fraction $f \ll 1$ that is white and uncorrelated; the precise value of f does not affect the results and is fixed to $f = 0.01$ throughout the analysis. The square of the Euclidean distance between the centers of the bins $(\log m_{1,i}, \log m_{2,j})$ and $(\log m_{1,k}, \log m_{2,l})$ is denoted $\Delta^2 \log m_{ijkl}$. At zero separation, the variance is σ^2. The correlation length scale is set by the parameter τ. Lastly, δ_{ijkl} is the Kroniker delta function.

In summary, the GP parameter set Γ consists of the free parameters μ, σ, and τ, and the fixed parameters μ_{ij} and f. For the prior on these hyper-parameters (written $p(\Gamma)$ in Eq. 6.23), we take a broad Gaussian prior for μ with mean 0 and standard deviation 10 and a half-Gaussian prior on $\sigma > 0$ with mean 0 and standard deviation 1.

We take a Gaussian prior for $\log \tau$. The mean and standard deviation of the Gaussian are chosen to place the width of the smallest mass bin 2-σ below the mean and the width of the mass range considered ($3\,M_\odot$ to $150\,M_\odot$) 2-σ above the mean. Thus the prior constrains the correlation length of the rate in log-mass to be typically longer than one bin, but shorter than the entire mass range. The correlation length τ is poorly constrained by the data with only ten detections, and the recovered posterior is very similar to the prior. We stress that in the limit of a large number of detections, the likelihood will dominate the GP prior and the posterior on the bin heights will become independent of these prior choices.

To evaluate the posterior of Eq. 6.23 we also need the expected number of detections λ_{ij} in bin ij. To do this we evaluate the sensitive spacetime volume in the ij-th bin, $\langle VT \rangle_{ij}$, so that:

$$\lambda_{ij} = \mathcal{R}_{ij} \langle VT \rangle_{ij}, \tag{6.26}$$

The sensitive volume was introduced in Section 6.4, and is calculated from the detection probability $P_{\rm det}$ by [19]:

$$VT(m_1, m_2) = T_{\rm obs} \int P_{\rm det}(m_1, m_2, z) \frac{1}{1+z} \frac{{\rm d}V_c}{{\rm d}z} {\rm d}z, \tag{6.27}$$

where $T_{\rm obs}$ is the observing time. This assumes that the merger rate is constant (non-evolving with redshift) in comoving volume and source-frame time, and that the detector sensitivity is constant in time. The average sensitive volume in bin ij is:

$$\langle VT \rangle_{ij} = \int_{m_{i-1/2}}^{m_{i+1/2}} \int_{m_{j-1/2}}^{m_{j+1/2}} VT(m_1, m_2) \frac{{\rm d}m_1}{m_1} \frac{{\rm d}m_2}{m_2}, \tag{6.28}$$

where the integral boundaries are the edges of the bin.

Consistent with the other analyses in this work, we calculate $P_{\rm det}(m_1, m_2, z)$ under the assumption of a single-detector SNR threshold. [19] found this approximation to $VT(m_1, m_2)$ to overestimate the sensitive volume by a factor of ~ 1.6 relative to the volume calculated by injections into the detection pipeline. This factor of ~ 1.6 between the semi-analytic sensitive volume and the injection-estimated volume is fairly constant across the mass space. We therefore divide our value of $\langle VT \rangle_{ij}$ by 1.6 to match the results of injections. We further assume that the total $\langle VT \rangle$ has a (Gaussian) 1-σ error of 18% (arising from an amplitude calibration error of $\sim 6\%$) and marginalize over this uncertainty, split evenly between the mass bins [3].

CHAPTER 7

A STANDARD SIREN MEASUREMENT OF THE HUBBLE CONSTANT FROM GW170817 WITHOUT THE ELECTROMAGNETIC COUNTERPART

The previous chapters focused on analyzing the masses, spins, and distances of binary black holes in order to learn about the astrophysical processes by which they form and merge. In this chapter, we turn to another type of measurement enabled by gravitational-wave detections: the cosmological expansion of the universe.

Gravitational waves from a compact binary merger provide a direct measurement of the distance to the source, lending them the name "standard sirens." The basic reason for this is as follows. The amplitude of a gravitational-wave signal decays inversely with the luminosity distance from the source. Meanwhile, the intrinsic "loudness" of the gravitational-wave signal is set by the binary's chirp mass; this same chirp mass sets the leading-order frequency evolution of the waveform. This means that by measuring the frequency evolution, we can extract the intrinsic loudness of the source, and thereby work out the distance to the source.

In order to measure the expansion of the universe and infer cosmological parameters, we need to measure the redshift-distance relationship. For a given source, the gravitational-wave signal provides its luminosity distance, but we need an independent measurement of redshift. The most straightforward approach is to observe the same source with gravitational waves and light, inferring distance from the gravitational-waves and redshift from the electromagnetic emission. An electromagnetic counterpart to a gravitational-wave source has so far been observed only once, with the first binary neutron star merger GW170817. In practice, we cannot measure the redshift of the electromagnetic counterpart directly, but we use the electromagnetic counterpart to identify a unique host galaxy, from which we can

easily obtain a redshift. Combining the gravitational-wave measurement of the distance to GW170817 with the redshift of its host galaxy, NGC 4993, yielded the first standard siren measurement of the Hubble constant [9]. However, most gravitational-wave sources do not have electromagnetic counterparts. In fact, Bernie Schutz's original proposal to measure the Hubble constant with gravitational waves did not rely on an electromagnetic counterpart and unique host galaxy, but rather a catalog of potential host galaxies. The galaxy catalog provides a statistical distribution of possible redshifts, and by stacking the measurement from multiple gravitational-wave observations, we expect to converge to the correct Hubble constant value.

In this work, as a proof of concept, we apply the galaxy-catalog based method to measure the Hubble constant with GW170817. This represents the first application of the galaxy-catalog standard siren method to real data. Because GW170817 was a relatively nearby and well-localized gravitational-wave source, using a full galaxy catalog yields a Hubble constant measurement that is only ~ 2 times less informative than the measurement we obtained from the electromagnetic counterpart. Our simulations suggest that in most cases, the galaxy-catalog approach will provide a measurement of the Hubble constant that is ~ 3 times less informative than the result provided by a unique host galaxy. This work was carried out in collaboration with Daniel Holz and other members of the LIGO and Virgo collaborations, and published in the Astrophysical Journal Letters [104].

We thank Chihway Chang, Alison Coil and Risa Wechsler for valuable discussions about galaxy properties. MF was supported by the NSF Graduate Research Fellowship Program under grant DGE-1746045. MF and DEH were supported by NSF grant PHY-1708081. They were also supported by the Kavli Institute for Cosmological Physics at the University of Chicago through NSF grant PHY-1125897 and an endowment from the Kavli Foundation. DEH also gratefully acknowledges support from a Marion and Stuart Rice Award. The authors acknowledge the Italian Istituto Nazionale di Fisica Nucleare (INFN), the French

Centre National de la Recherche Scientifique (CNRS) and the Foundation for Fundamental Research on Matter supported by the Netherlands Organisation for Scientific Research, for the construction and operation of the Virgo detector and the creation and support of the EGO consortium. We are grateful to the LIGO and Virgo collaborations for releasing sky map data for GW170817 at dcc.ligo.org/LIGO- P1800061/public. This work has made use of CosmoHub. CosmoHub has been developed by the Port d'Informaci Cientfica (PIC), maintained through a collaboration of the Institut de Fsica d'Altes Energies (IFAE) and the Centro de Investigaciones Energticas, Medioambientales y Tecnolgicas (CIEMAT), and was partially funded by the "Plan Estatal de Investigacin Cientfica y Tcnica y de Innovacin" program of the Spanish government. We also acknowledge the First Two Years data release (https://www.ligo.org/scientists/first2years/).

7.1 Abstract

We perform a statistical standard siren analysis of GW170817. Our analysis does not utilize knowledge of NGC 4993 as the unique host galaxy of the optical counterpart to GW170817. Instead, we consider each galaxy within the GW170817 localization region as a potential host; combining the redshift from each galaxy with the distance estimate from GW170817 provides an estimate of the Hubble constant, H_0. We then combine the H_0 values from all the galaxies to provide a final measurement of H_0. We explore the dependence of our results on the thresholds by which galaxies are included in our sample, as well as the impact of weighting the galaxies by stellar mass and star-formation rate. Considering all galaxies brighter than $0.01L_B^\star$ as equally likely to host a BNS merger, we find $H_0 = 76^{+48}_{-23}$ km s^{-1} Mpc^{-1} (maximum *a posteriori* and 68.3% highest density posterior interval; assuming a flat H_0 prior in the range $[10, 220]$ km s^{-1} Mpc^{-1}). Restricting only to galaxies brighter than $0.626L_B^\star$ tightens the measurement to $H_0 = 77^{+37}_{-18}$ km s^{-1} Mpc^{-1}. We show that weighting the host galaxies by stellar mass or star-formation rate provides entirely consistent results

with potentially tighter constraints. While these statistical estimates are inferior to the value from the counterpart standard siren measurement utilizing NGC 4993 as the unique host, $H_0 = 76^{+19}_{-13}$ km s^{-1} Mpc^{-1} (determined from the same publicly available data), our analysis is a proof-of-principle demonstration of the statistical approach first proposed by Bernard Schutz over 30 years ago.

7.2 Introduction

The first multi-messenger detection of a binary neutron star (BNS) merger, GW170817, by LIGO [1] and Virgo [23] enabled the first standard siren measurement of the Hubble constant, H_0, ushering in the era of gravitational-wave (GW) cosmology [9, 14, 15]. This H_0 measurement combined the luminosity distance to the source, as measured from the GW signal [219], with the known redshift of the host galaxy, NGC 4993. NGC 4993 was identified as the unique host galaxy following the discovery of an optical transient located only ~ 10 arcsec from NGC 4993 [74, 229, 15]. The probability of a chance coincidence between the GW signal and the optical transient was estimated to be $\lesssim 0.5\%$ [229], and the probability of a chance association between the optical transient and NGC 4993 is $\lesssim 0.004\%$ [9].

The original proposal by [219] to measure the Hubble constant with GW detections of compact binary mergers did not involve electromagnetic counterparts. Instead, Schutz considered bright galaxies in the GW localization region as potential hosts to the merger. Each galaxy provides a redshift that, when combined with the GW-measured luminosity distance, gives a separate estimate of H_0. The final H_0 measurement from a single event is the sum of all contributions from the individual galaxies. The first detailed exploration of this method on simulated data, and with the first use of a galaxy catalog [25], was by [81]. An up-to-date forecast incorporating realistic detection rates, galaxy peculiar velocities, large-scale structure, and additional considerations can be found in [70]. We refer to this approach of measuring H_0 as the "statistical" method [219, 172, 205, 70], compared with

the "counterpart" method in which an electromagnetic (EM) counterpart provides a unique host galaxy association. In the limit where the GW event is so well-localized that there is only one potential host galaxy in the GW localization error box [68], the statistical method reduces to the counterpart method. In the opposite limit, where the GW event is poorly localized, there are so many potential host galaxies that the distinct peaks from individual galaxies are washed out, and the H_0 measurement is uninformative [70].

The statistical approach may be the only way to do standard siren science with binary black holes, since they are not expected to have EM counterparts. We emphasize that although the statistical measurements for a given event are inferior to the counterpart case, combining many of these measurements leads to increasingly precise constraints [219, 81, 70, 193]. In ground-based gravitational wave detector networks, the rate of detection of binary black holes is significantly higher than that for neutron stars [3, 11, 14],although the higher rate is not expected to compensate for the inferior constraints [70]. Nonetheless, the black hole systems can be observed to much higher redshifts, potentially providing constraints on the evolution history of the Universe out past the turnover between dark matter and dark energy domination [81, 91, 44, 109]. Because these systems are farther away, however, it will be a greater challenge to supply a sufficiently complete galaxy catalog.

In this paper we carry out a measurement of H_0 using the GW data from GW170817 and a catalog of potential host galaxies within the GW localization region. In other words, we explore how tight the H_0 measurement from GW170817 would have been if an EM counterpart had not been detected or if a unique host galaxy had not been identified. We present our methods in §7.3, a discussion of the galaxy selection in §7.4, a discussion of the gravitational-wave constraints in §7.5, results in §7.6, and conclude in §7.7.

7.3 Methods

We follow the statistical framework presented in [70] [see also 81, 124]. We include the role of GW selection effects, galaxy catalog incompleteness, galaxy luminosities, and redshift uncertainties in our analysis. The posterior on H_0 given the GW and EM data, $x_{\rm GW}$ and $x_{\rm EM}$, is:

$$p(H_0 \mid x_{\rm GW}, x_{\rm EM}) = \frac{p_0(H_0)}{\beta(H_0)} \int p(x_{\rm GW}|\hat{D}_L(z, H_0), \Omega) p(x_{\rm EM}|z, \Omega) p_0(z, \Omega) d\Omega dz, \qquad (7.1)$$

where $\hat{D}_L(z, H_0)$ is the luminosity distance of a source at redshift z for a given H_0 [fixing other cosmological parameters to the Planck values; 24][1], Ω is the sky position, and $\beta(H_0)$ is a normalization term to ensure that the likelihood normalizes to 1 when integrated over all *detectable* GW and EM datasets [179]. The term $p_0(H_0)$ represents the prior on the Hubble constant. A detailed derivation of Equation 7.1, including the role of the normalization term $\beta(H_0)$, is provided in the Appendix.

As first emphasized by [219], the GW signal from a compact binary coalescence allows for a direct measurement of the distance to the source, as well as its sky location. This measurement is represented in the GW likelihood term, $p(x_{\rm GW}|D_L, \Omega)$, which is the probability of the GW data in the presence of signal from a compact binary with parameters D_L and Ω marginalized over the other parameters of the signal (including the inclination angle, component masses, spins and/or tides). The corresponding posterior $p(D_L, \Omega \mid x_{\rm GW}) \propto p(x_{\rm GW}|D_L, \Omega) p_0(D_L, \Omega)$ is summarized in the 3-dimensional sky map, which provides a fit to the posterior samples provided by the GW parameter estimation pipeline LALInference [245, 223, 224]. For this analysis, we use the publicly released 3-dimensional sky map from [20] (see §7.5 and §7.6). To get the likelihood from the posterior probability, we must first divide out the default "volumetric" distance prior, $p_0(D_L, \Omega) \propto D_L^2$.

1. For the redshifts considered here, $z \lesssim 0.05$, other cosmological parameters affect the distance-redshift relation at the sub-percent level, and so our analysis is insensitive to their precise values.

Meanwhile, the EM likelihood term $p(x_{\rm EM} \mid z, \Omega)$ is the probability of the electromagnetic data in the presence of signal from a compact binary with parameters z and Ω. In the absence of an EM counterpart and/or a host galaxy identification, we assume the measurement $p(x_{\rm EM}|z, \Omega)$ is completely uninformative, and set:

$$p(x_{\rm EM}|z, \Omega) \propto 1. \tag{7.2}$$

In this case, the redshift information enters only through the prior term, $p_0(z, \Omega)$, which we take to be a galaxy catalog. The detection of an electromagnetic counterpart typically results in $p(x_{\rm EM} \mid z, \Omega)$ being strongly peaked around some $\hat{\Omega}$ allowing the identification of a host galaxy. We note that in some cases an optical transient may be identified, but it may not be possible to uniquely identify the associated host galaxy. In these circumstances one could perform a pencil-beam survey of the region surrounding the transient (e.g., at distances of $\lesssim 100\,{\rm kpc}$ from the line-of-sight to the transient), and sharply reduce the relevant localization volume [70]. This reduces the number of potential host galaxies, and thereby improves the measurement.

The galaxy catalog term $p_0(z, \Omega)$ is given by:

$$p_0(z, \Omega) = f p_{\rm cat}(z, \Omega) + (1 - f) p_{\rm miss}(z, \Omega), \tag{7.3}$$

where $p_{\rm cat}$ is a catalog of known galaxies, $p_{\rm miss}$ represents the distribution of missing galaxies, and f denotes the overall completeness fraction of the catalog. The contribution from the known galaxies is:

$$p_{\rm cat}(z, \Omega) = \sum_i^{N_{\rm gal}} w_i N(\bar{z}_i, \sigma_z; z) N(\bar{\Omega}_i, \sigma_\Omega; \Omega), \tag{7.4}$$

where $\bar{z}_i, \bar{\Omega}_i$ denotes the (peculiar velocity-corrected) "Hubble" redshifts and sky coordinates of all galaxies in the catalog, and $N(\mu, \sigma; x)$ denotes the normal probability density function

with mean μ and standard deviation σ evaluated at x. To account for peculiar velocity uncertainties, which can be significant for nearby sources, we assume that the true Hubble velocity is normally distributed about the measured Hubble velocity with an uncertainty of $c\sigma_z$ [220]. On the other hand, the uncertainty on the sky coordinates of galaxies in the catalog is negligible for our purposes, so we always approximate $N(\bar{\Omega}, \sigma_\Omega; \Omega)$ by $\delta(\bar{\Omega} - \Omega)$.

The weights w_i can be chosen to reflect the *a priori* belief that a galaxy could host a gravitational-wave source. For example, setting all weights to $w_i = \frac{1}{N_{\rm gal}}$ corresponds to equal probability for each galaxy to host a gravitational wave source. In general, since we might expect that the BNS rate is traced by some combination of stellar mass and/or star formation rate [206, 167, 110, 72], we may assign unequal weights to galaxies based on these (or any other relevant observable) quantities, ensuring that the weights sum to unity. In the following, we use a galaxy's B-band luminosity as a proxy for its star formation rate, and its K-band luminosity as a proxy for its total stellar mass [46, 223]; these are very rough estimates, but serve to demonstrate the method. In these cases, we apply weights proportional to B-band or K-band luminosity, $w_i \propto L^i_B$ or $w_i \propto L^i_K$, and explore the dependence of the result on these of weightings. We should probably have a discussion somewhere about bias due to GRBs/KN somehow preferentially selecting face-on galaxies?

To calculate the term $p_{\rm miss}$ in Equation 7.3, we assume that on large scales, the distribution of galaxies, $p_0(z, \Omega)$ is uniform in comoving volume. Let $p_{\rm vol}(z, \Omega)$ denote the cosmologically homogeneous and isotropic distribution normalized over the volume contained in the range $z_{\rm min} < z < z_{\rm max}$ considered in our analysis. (The result does not depend on the choice of $z_{\rm min}$ or $z_{\rm max}$ provided that the interval encompasses all possible redshifts of the source for all allowed values of H_0.) Assuming all galaxies are weighted equally, the distribution of missing galaxies is written as:

$$p_{\rm miss}(z, \Omega) = \frac{\left[1 - P_{\rm complete}(z)\right] p_{\rm vol}(z, \Omega)}{(1 - f)}, \tag{7.5}$$

where $P_{\rm complete}(z)$ is the probability that a galaxy at redshift z is in the catalog, and the completeness fraction f is given by:

$$f = \int_{z_{\rm min}}^{z_{\rm max}} P_{\rm complete}(z) p_{\rm vol}(z, \Omega)\, dz\, d\Omega. \tag{7.6}$$

We can similarly add galaxy weightings to an incomplete catalog by computing the luminosity distribution of the "missing galaxies" as a function of redshift, $p(L \mid z, \text{missing})$. We calculate this distribution by assuming that the luminosities of the missing galaxies together with those in the catalog are distributed according to the Schechter function. Then, the weights of the missing galaxies are given by:

$$w(z) \propto \int L p(L \mid z, \text{missing})\, dL, \tag{7.7}$$

and, weighting each missing galaxy by its luminosity, Equation 7.5 becomes:

$$p_{\rm miss}(z, \Omega) \propto w(z) p_{\rm miss}(z, \Omega). \tag{7.8}$$

Note that we have assumed that the coverage of the catalog is uniform over the sky and so $P_{\rm complete}$ is independent of Ω. (This is true over the relevant sky area for the present analysis, but the method can be easily generalized to add an Ω-dependence.) Alternate approaches of taking into account the incompleteness of galaxy catalogs are being explored in [124]. However, in the present case of a single nearby source where the catalog is largely complete, the differences in results from the various approaches are small, and in particular, well within the statistical uncertainties. In section 7.4, the completeness function $P_{\rm complete}$ is estimated for the galaxy catalog used in the analysis.

To demonstrate the statistical method, we apply the analysis described above to 249 simulated BNS GW detections from the First Two Years (F2Y) catalog [225] and the MICE

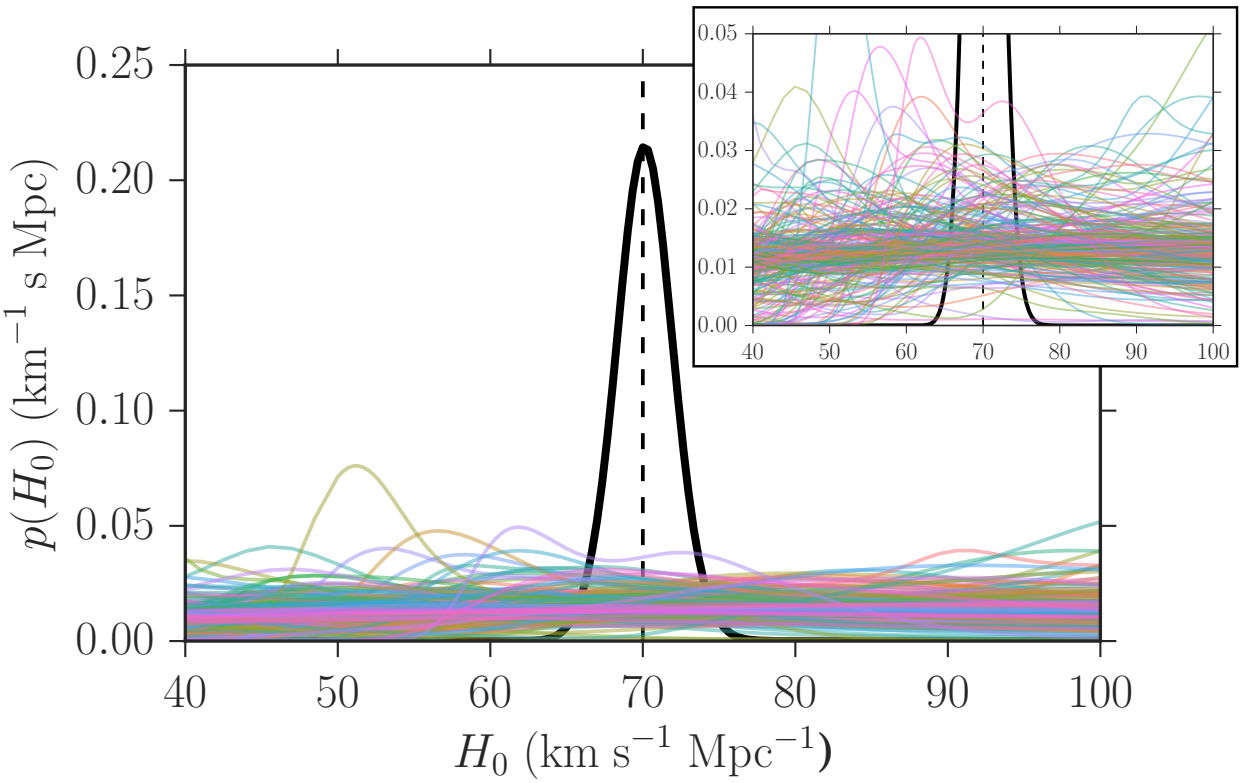

Figure 7.1: **Projected H_0 constraints** using the statistical method on a sample of 249 simulated BNS detections and the MICE mock galaxy catalog. The thin colored lines show the H_0 posteriors from individual events, while the solid black curve shows the combined posterior. The prior is assumed to be flat in all cases. The dashed black line shows the injected value, $H_0 = 70$ km s^{-1} Mpc^{-1}.

simulated galaxy catalog [113, 75, 114, 61, 130, 62]. We assign each BNS detection from the F2Y 2016 scenario (roughly corresponding to O2) to a galaxy in the MICE catalog with a redshift that matches the injected distance and assumed H_0 value ($H_0 = 70$ km s^{-1} Mpc^{-1}). For each event, we rotate the sky coordinates of the galaxies in the catalog so that the sky position of the host galaxy matches the true sky position of the BNS injection. We then carry out the statistical method using the 3-dimensional sky map for each mock BNS and the galaxies in MICE, assuming no peculiar velocities or incompleteness, and assigning weights to the galaxies in MICE so that the redshift distribution matches the injected redshift distribution of the F2Y dataset, $p(z) \propto z^2$. This last step is necessary in order to ensure that the selection effects are incorporated consistently between the injections and the likelihood. The results are shown in Fig. 7.1. Even in the best-case scenario of perfectly-known galaxy redshifts and a complete catalog, the H_0 posteriors from most individual events are nearly flat over the prior range. Combining the 249 individual events, the final H_0 posterior is $H_0 = 70.1^{+1.9}_{-1.9}$ km s^{-1} Mpc^{-1} (68.3% credible interval), corresponding to a convergence rate of $\sim 40\%/\sqrt{N}$, consistent with [70]. As is visible in the Figure, we confirm that the method is unbiased, with the result for large numbers of detections approaching the true value of H_0. We note that most of the simulated detections in the F2Y dataset have much larger localization volumes than GW170817, which was an unusually loud, nearby source that was detected while all three detectors were operational. Therefore, we expect the statistical H_0 measurement from GW170817 to be unusually informative compared to an average event. We quantify this expectation in Section 7.6.

7.4 Galaxy catalogs

To measure H_0 statistically with GW170817, we use version 2.3 of the GLADE galaxy catalog to construct our redshift prior in Equation 7.3 [78]. GLADE provides galaxy redshifts in the heliocentric frame, corrected for peculiar motions using the peculiar velocity catalog

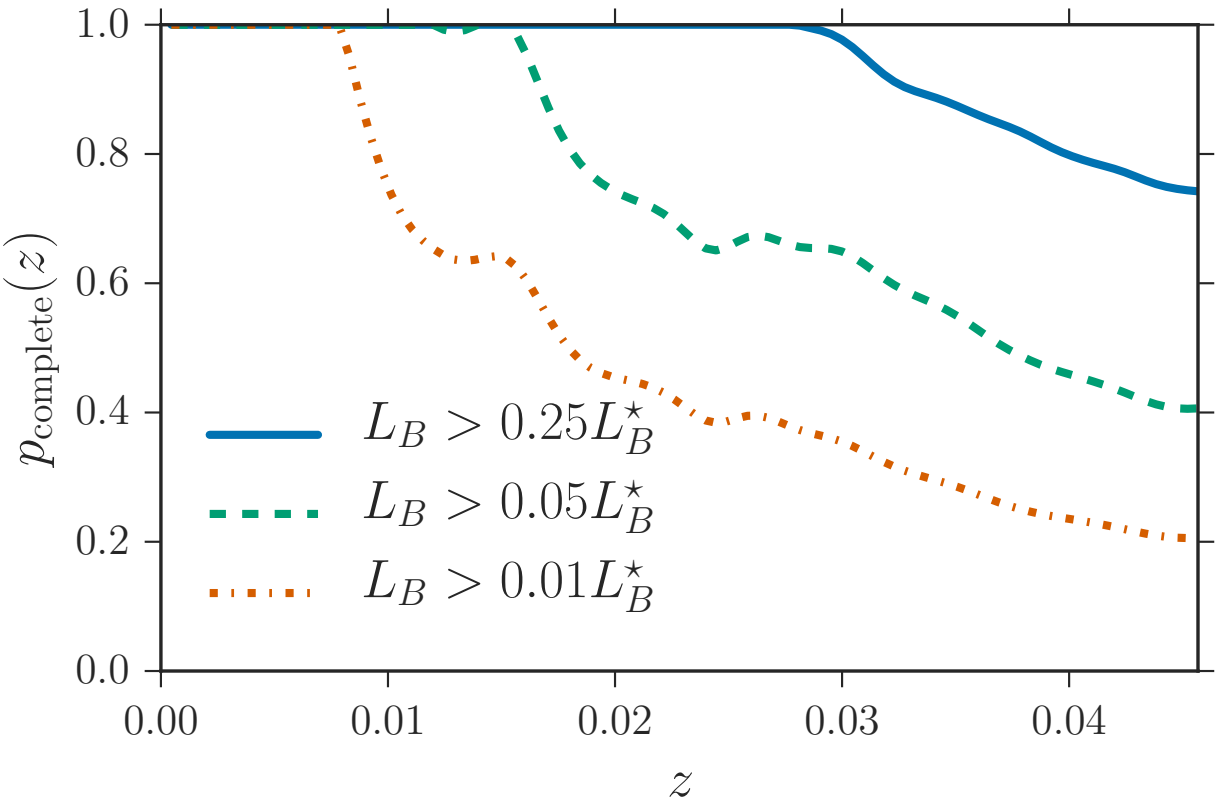

Figure 7.2: **Completeness of the GLADE catalog** as a function of redshift for galaxies brighter than $0.25L_B^\star$ (solid blue curve), $0.05L_B^\star$ (dashed green curve), and $0.01L_B^\star$ (dot-dahsed orange curve), calculated by comparing the redshift distribution of galaxies in GLADE to a distribution that is constant in comoving volume. For galaxies brighter than $0.626L_B^\star$, GLADE is complete across the entire redshift range shown.

of [63]. For galaxies which are also listed in the group catalog of [156], as identified by a common Principal Galaxy Catalog (PGC) identifier, we apply an additional correction to correct their velocities to the radial velocity of the group. We assume the group velocity is given by the unweighted mean of the velocities of all member galaxies, although we note that for the dominant group containing NGC 4993, careful group modeling has been done [129]. Finally, we correct all heliocentric velocities to the reference frame of the cosmic microwave background [128] and assign a 200 km/s Gaussian uncertainty to the "Hubble velocity" of each galaxy in the catalog [corrected by all peculiar motions; 63, 220].

GLADE also provides luminosity information for galaxies, listing apparent magnitudes in the B-, J-, H-, and K-bands. We use the reported B-band luminosities to characterize the completeness of the catalog (a small fraction of galaxies do not have B-band apparent magnitudes reported in the catalog; we remove these galaxies from our analysis, assuming that their magnitudes are below our adopted luminosity cutoff). Following [119] and [31], we adopt B-band Schechter function parameters $\phi^* = 5.5 \times 10^{-3}\ h_{0.7}^3\ \mathrm{Mpc}^3$, $\alpha_B = -1.07$, $L_B^* = 2.45 \times 10^{10}\ h_{0.7}^2 L_{B,\odot}$ throughout. The corresponding characteristic absolute magnitude is $M_B^* = -20.47 + 5\log_{10} h_{0.7}$. We will also consider the K-band magnitudes reported in GLADE when applying galaxy weights, and we use the K-band Schechter function parameters of $\alpha_K = -1.02$, $M_K^* = -23.55 + 5\log_{10} h_{0.7}$ [171].

Figure 7.2 summarizes the completeness of GLADE as a function of redshift. We find that GLADE is complete up to redshifts $z \sim 0.06$ for galaxies brighter than $\sim 0.626 h_{0.7}^{-2}\ L_B^\star$, corresponding to about 0.66 of the Milky Way luminosity for $h_{0.7} = 1$. Galaxies brighter than $\sim 0.626 L_B^\star$ make up half of the total luminosity for the given Schechter function parameters. We find that for $z \lesssim 0.03$, GLADE is complete for galaxies down to 2.5 times dimmer, or $\sim 0.25 L_B^\star$, corresponding to $M_B = -18.96 + 5\log_{10} h_{0.7}$ [see also Figure 2 of 31]. Such galaxies make up 75% of the total B-band luminosity. If we consider galaxies down to $\sim 0.05 L_B^\star$ ($M_B = -17.22 + 5\log_{10} h_{0.7}$), GLADE is $\sim$ 70% complete at $z \sim 0.03$, and

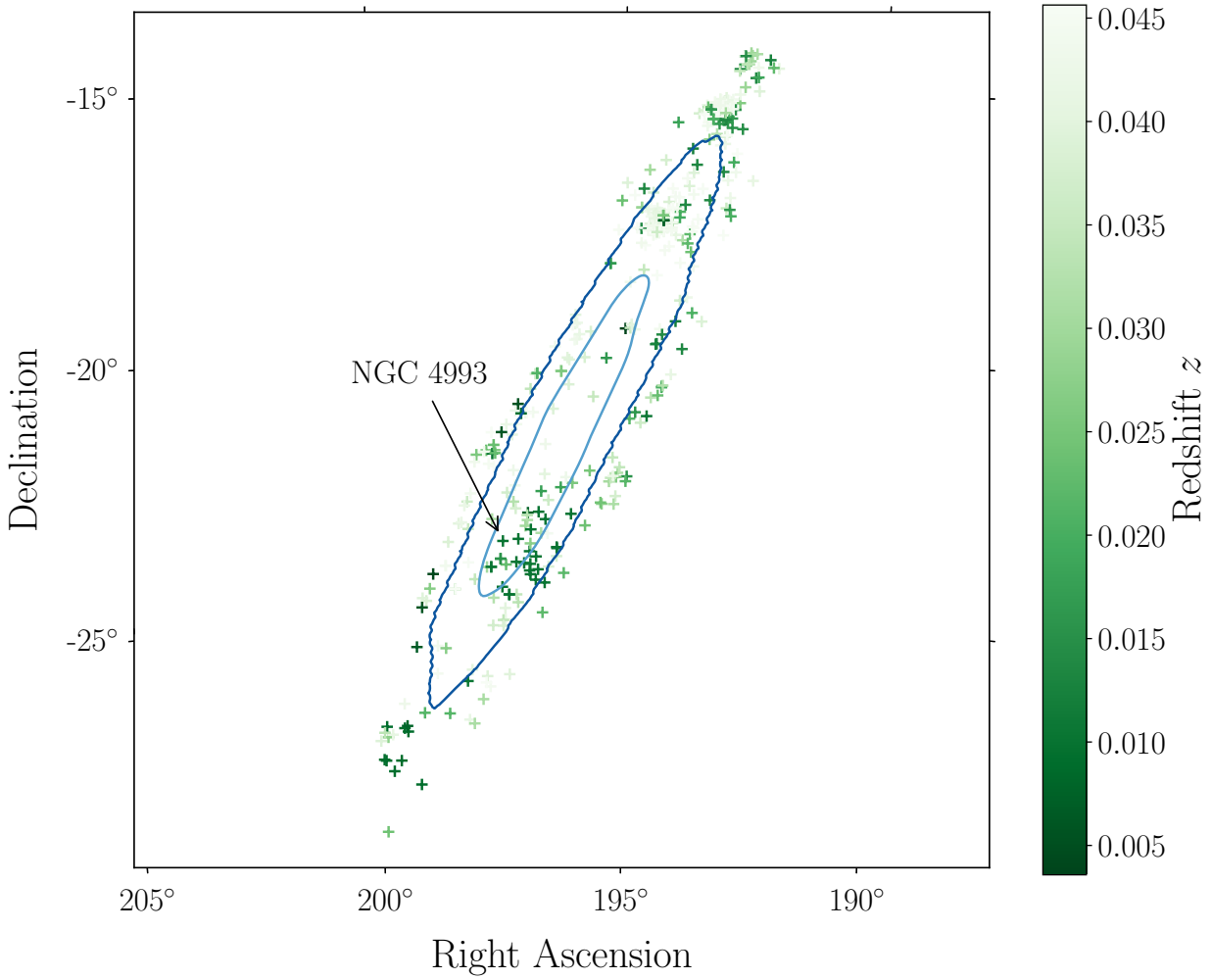

Figure 7.3: **Two-dimensional localization region** of GW170817 (blue contours) with the sky coordinates of the 408 GLADE galaxies (green crosses) within the 99% localization area and the redshift range $0 < z \lesssim 0.046$ (for an H_0 prior range of $H_0 \in [10, 220]$ km s^{-1} Mpc^{-1}). The light and dark blue contours enclose the 50% and 90% probability regions, respectively, and the shading of the galaxy markers denotes their redshifts, corrected for peculiar and virial motions as described in the text.

even if we consider galaxies down to $\sim 0.01 L_B^\star$ ($M_B = -15.47 + 5\log_{10} h_{0.7}$), including 99% of the total B-band luminosity, GLADE is $\gtrsim 80\%$ complete for $z \lesssim 0.01$, and $\sim 40\%$ complete at $z \sim 0.03$. In the K-band, we find that with our assumed K-band Schechter function parameters, GLADE is complete up to $z \sim 0.045$ for galaxies with $L_K > 0.36 L_K^\star$, which contain 70% of the total K-band luminosity, and up to $z \sim 0.03$ for galaxies with $L_K > 0.1 L_K^\star$, which contain 90% of the total luminosity. For galaxies brighter than $L_K = 0.005 L_K^\star$, which make up more than 99% of the total K-band luminosity, GLADE is $\sim 70\%$ complete at $z = 0.01$.

7.5 Source Localization and Distance

From the GW data alone, GW170817 is the best-localized GW event to date. The original analysis by the LIGO-Virgo collaboration [14] reported a 90% localization area of 28 deg^2 and a 90% localization volume of 380 Mpc^3 [assuming Planck cosmology; 24], while the most recent analysis [20] improves the 90% localization area to 16 deg^2 and the 90% volume to 215 Mpc^3. We use this updated 3-dimensional sky map [223, 224] from [20] throughout.[2] Figure 7.3 shows the 2-dimensional sky map together with the galaxies in the GLADE catalog within the localization region. Figure 7.4 shows that, although there are a total of 408 galaxies within the 99% localization area (see Figure 7.3), most of the galaxies with high sky-map probability come in a few distinct groups: a dominant group at $z \sim 0.01$ regardless of the assumed luminosity threshold, followed by a secondary group at $z \sim 0.006$ containing only moderately faint galaxies. Therefore, there are only a few distinct redshifts that can possibly correspond to the measured distance of GW170817, and we expect that combining the galaxy catalog with the GW localization will yield an informative measurement of the Hubble constant.

2. With the data release accompanying [20], the LIGO-Virgo collaboration has made the 3-dimensional data behind this sky map publicly available at the following url: https://dcc.ligo.org/DocDB/0150/P1800061/009/figure_3.tar.gz

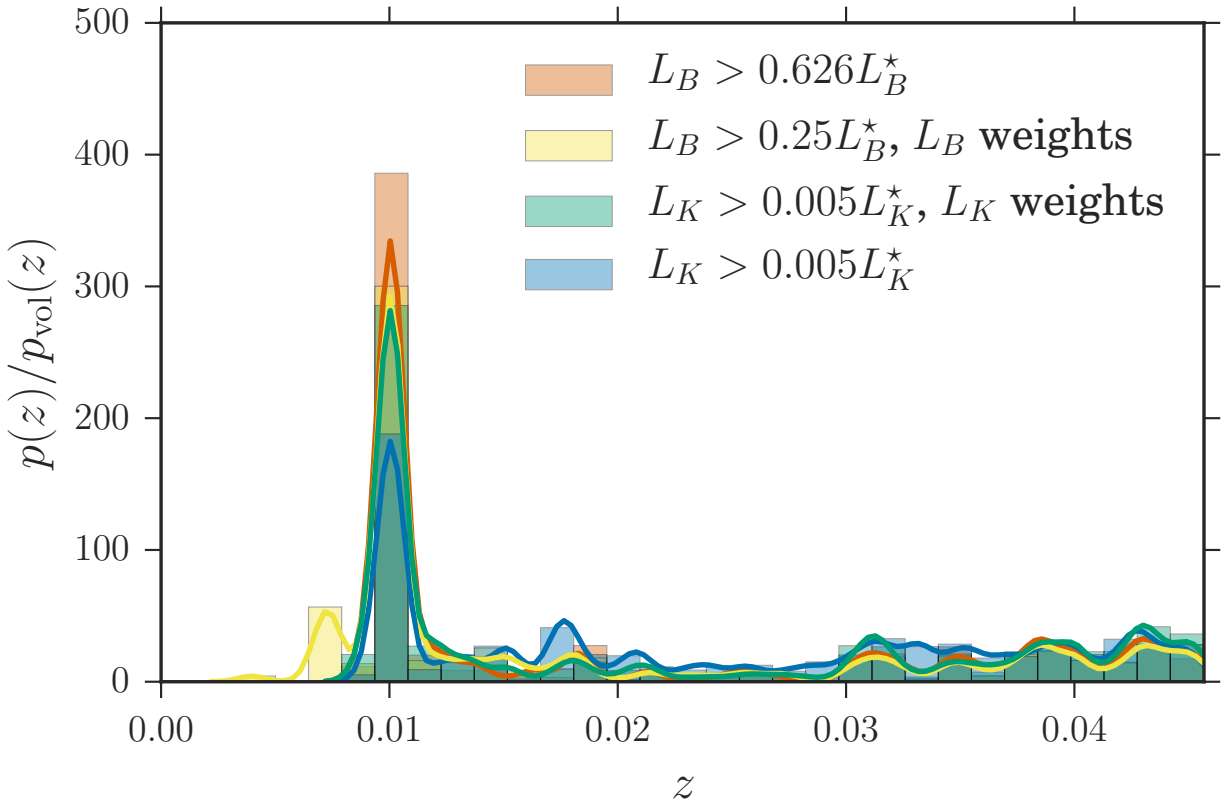

Figure 7.4: **Probability distribution of the redshifts of potential hosts to GW170817** weighted by the GW sky map probability, $p(z) = \int p(x_{\rm GW} \mid \Omega) p_0(z, \Omega) d\Omega$, compared to a uniform in comoving volume distribution of galaxies, $p_{\rm vol}(z)$. For the orange histogram, we include all galaxies in the catalog brighter than $0.626L_B^\star$. For galaxies brighter than $0.626L_B^\star$, the catalog is complete over the redshift range. However, when we lower the luminosity cutoff to $0.25L_B^\star$ (yellow histogram) or $0.005L_K^\star$ (green and blue), we must account for catalog incompleteness at higher redshifts by considering the redshift and luminosity distributions of the missing galaxies (see §7.3). The yellow (green) histogram additionally weights each galaxy by its B-band (K-band) luminosity. If the ratio $p(z)/\ p_{\rm vol}\,(z)$ were completely flat, we would expect an uninformative H_0 measurement in which our posterior recovers our prior. However, in all instances there is a dominant peak at $z \sim 0.01$, suggesting that the resulting H_0 measurement will be informative. Adding in luminosity weights, especially in the K-band, makes the peak more dominant.

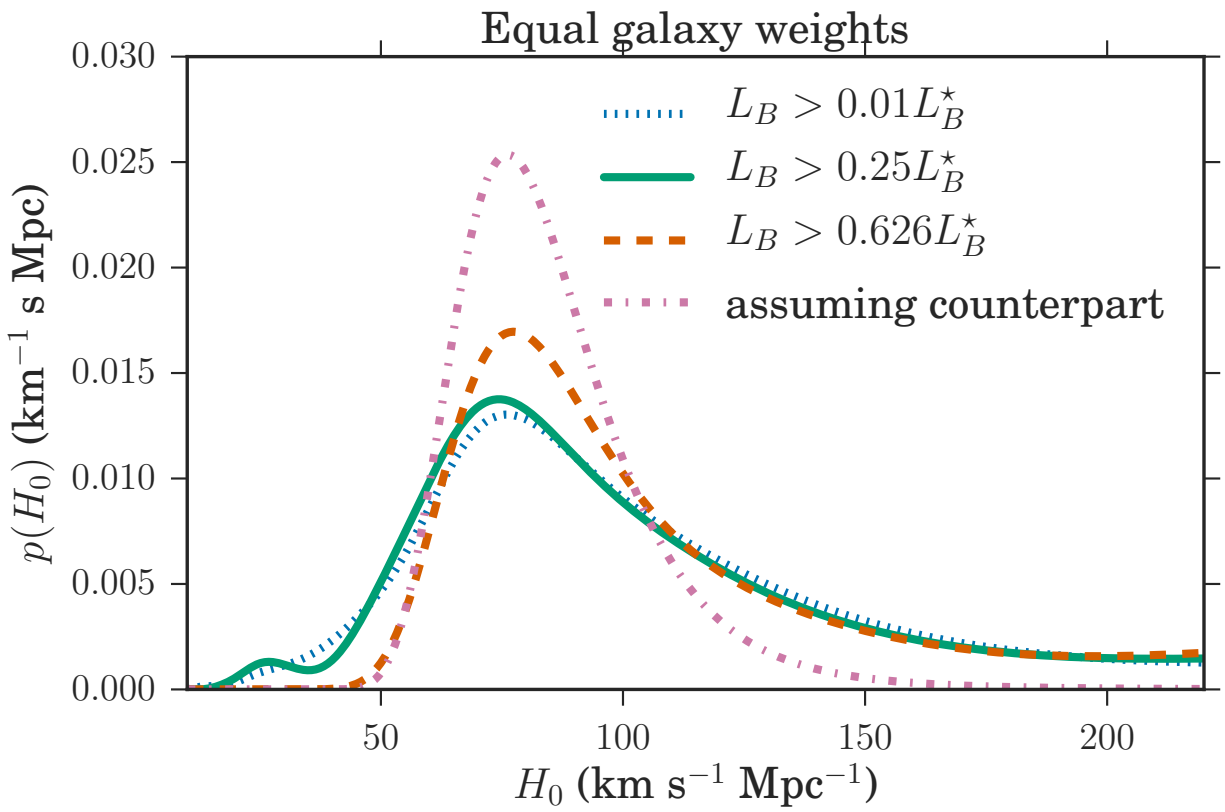

Figure 7.5: **Posterior probability of** H_0 under various assumptions regarding the potential host galaxy. We adopt a flat H_0 prior in the range $H_0 \in [10, 220]$ km s^{-1} Mpc^{-1}. For the dashed orange curve, we assume that only galaxies brighter than $0.626L_B^\star$ (containing 50% of the total luminosity) can host BNS events, meaning that the galaxy catalog is complete over the relevant redshift range. The solid green curve lowers the luminosity cutoff to $0.25L_B^\star$ (containing 75% of the total luminosity), and accounts for the mild incompleteness of the catalog above redshifts $z \sim 0.03$. The dotted blue curves incorporate all galaxies brighter than $0.01L^\star$ (containing 99% of the total luminosity), accounting for the incompleteness of faint galaxies at redshifts $z \gtrsim 0.01$. The dot-dashed pink curve shows the H_0 measurement assuming the host galaxy is known to be NGC 4993.

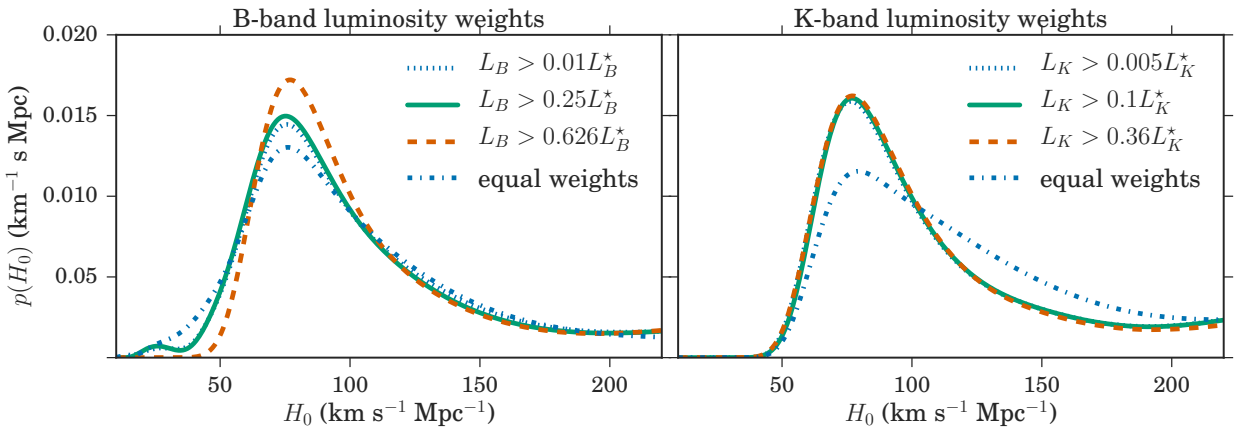

Figure 7.6: **Posterior probability of H_0, weighting all galaxies in the volume by their B-band luminosities**, corresponding roughly to weighting by star-formation rate (left), or K-band luminosities, corresponding roughly to weighting by stellar mass (right). We have applied the necessary completeness correction (see §7.3). The blue dashed-dot curve shows all galaxies brighter than $0.01L_B^\star$ in B-band (left) or $0.005L_K^\star$ in K-band (right) with equal weights for comparison. Weighting galaxies by their K-band luminosities brings all the curves into very close agreement, because many galaxies in the group at $z \sim 0.01$ have brighter than average K-band luminosities (brighter than $1.5L_K^\star$) and thus dominate the K-band weighted population and contain the majority of the stellar mass.

The 3-dimensional sky map also provides an approximation to the luminosity distance posterior along each line-of-sight. As usual, the distance to GW170817 is determined directly from the gravitational waves, and is calibrated by general relativity [219]. No distance ladder is required.

7.6 Results

We combine the GW distance posterior for GW170817 with the redshift for each potential host galaxy within the localization region. As detailed in §7.3, each galaxy produces a posterior probability for H_0, and we combine these estimates among all the galaxies in the localization region to arrive at a final estimate for H_0. We adopt a flat prior in H_0 over the range 10–220 km s^{-1} Mpc^{-1}. The results are presented in Figure 7.5. Because the galaxies are predominantly found in one galaxy group at $z \sim 0.01$, the H_0 posterior shows a clear peak

at $H_0 \approx 76\,\mathrm{km\ s^{-1}\ Mpc^{-1}}$. And because NGC 4993, the true galaxy host of GW170817, is a member of the group at $z \sim 0.01$, we should not be surprised to learn that the peak in H_0 is consistent with the H_0 estimate from the GW170817 standard siren measurement including the counterpart [9]. Because this analysis has been performed on a 3-dimensional sky map using an approximation to the distance posteriors, rather than using the full 3-dimensional LIGO/Virgo posteriors, the results do not agree precisely with those of [9], and in particular, the position of the peak in Figure 7.5 is at $H_0 = 76\,\mathrm{km\ s^{-1}\ Mpc^{-1}}$ instead of $H_0 = 70\,\mathrm{km\ s^{-1}\ Mpc^{-1}}$. This is because our 3-dimensional sky map approximates the distance posterior along each line-of-sight by a simple 2-parameter Gaussian fit [see Eq. 1 of 223], which is an imperfect approximation to the true, asymmetric distance posterior [69, 82]. On the other hand, the analysis in [9] utilizes the full distance posterior along the line-of-sight to NGC 4993 rather than the Gaussian approximation.

Figure 7.5 shows four different posterior probability distributions, each using a different threshold for the galaxy catalog. In the "assuming counterpart" case, NGC 4993 (which is assumed to be the true host galaxy to GW170817) is given a weight of 1, and all the other galaxies in the localization volume are given a weight of 0. We find $H_0 = 76^{+19}_{-13}\,\mathrm{km\ s^{-1}\ Mpc^{-1}}$ (maximum *a posteriori* and 68.3% highest density posterior interval) for our default flat prior, or $H_0 = 74^{+18}_{-12}\,\mathrm{km\ s^{-1}\ Mpc^{-1}}$ for a flat-in-log prior (the prior choice in [9]). This peak is slightly shifted compared to the result presented in [9], $H_0 = 70^{+12}_{-8}\,\mathrm{km\ s^{-1}\ Mpc^{-1}}$, due to the usage of the Gaussian fit to the distance posterior found in the 3-dimensional sky map as discussed above.

The other curves in Figure 7.5 assume different limiting thresholds for what constitutes a potential host galaxy. For a luminosity threshold of $L > 0.626L^{\star}_{B}$, we find $H_0 = 77^{+37}_{-18}$ km s^{-1} Mpc^{-1} [3]. As the threshold is lowered, additional galaxies fall into the sample, and the H_0 posterior is broadened. For a limiting B-band magnitude of $0.25L^{\star}_{B}$, we need to account

3. The upper limits of the 68.3% highest density posterior intervals that we report here are especially sensitive to the upper limit we consider for the H_0 prior, 220 km s^{-1} Mpc^{-1}

for the incompleteness of the galaxy catalog at redshifts $z \gtrsim 0.03$, and for $0.01L_B^*$, we need to account for the incompleteness at $z \gtrsim 0.01$, as described in §7.3. The incompleteness correction leads to a slight additional broadening of the H_0 posterior, but the clear peak at $H_0 \approx 76\,\mathrm{km\ s^{-1}\ Mpc^{-1}}$ remains: we find $H_0 = 74^{+45}_{-24}$ km s^{-1} Mpc^{-1} for a luminosity threshold of $L > 0.25L_B^\star$ $H_0 = 76^{+48}_{-23}$ km s^{-1} Mpc^{-1} for a luminosity threshold of $L > 0.01L_B^\star$. This peak is the result of the galaxy group at $z \sim 0.01$, of which NGC 4993 is a member.

The curves in Figure 7.6 weight each galaxy by its B-band luminosity (a proxy for its recent star formation history; right) or its K-band luminosity (a proxy for its stellar mass; left). The peak at $H_0 \approx 76$ km s^{-1} Mpc^{-1} becomes more pronounced when galaxies are weighted by their luminosity, as the group containing NGC 4993 consists of many bright, mostly red galaxies. If we assume that the probability of hosting a BNS merger is proportional to a galaxy's B-band luminosity, the posterior on H_0 tightens from $H_0 \in [53, 124]$ km s^{-1} Mpc^{-1} (68.3% highest density posterior interval) when applying equal weights to all galaxies brighter than 0.01 $L_B^\star$ to $H_0 \in [54, 120]$ km s^{-1} Mpc^{-1}. Applying K-band luminosity weights to galaxies brighter than 0.005 $L_K^\star$, the 68.3% posterior interval tightens from $H_0 \in [61, 137]$ km s^{-1} Mpc^{-1} to $H_0 \in [57, 118]$ km s^{-1} Mpc^{-1}. Although these results are suggestive that weighting by stellar-mass or star-formation rate may lead to faster convergence, the properties of BNS host galaxies are still uncertain, and it is impossible to establish this definitively with a single event. As the source sample increases it is expected to relate to some combination of these quantities, and incorporating these trends will lead to improvements in the statistical H_0 analysis.

In order to quantify the degree of information in the GW170817 H_0 posterior compared to an "average" event as expected from the F2Y dataset, we consider the difference in the Shannon entropy between the flat prior and the posterior [see Appendix; 221]. We compare this measure of information for the statistical GW170817 H_0 posterior to the individual

statistical H_0 posteriors from each of the simulated BNS events in Section 7.3. We find that for a flat prior in the (relatively narrow) range $H_0 \in [40, 100]$ km s^{-1} Mpc^{-1}, the information gained by applying the statistical method to GW170817 is 0.34 bits. Meanwhile, the median information in an individual posterior shown in Figure 7.1 is only 0.047 bits, so that GW170817 is in the top $\sim 90\%$ of informative events, even under optimistic assumptions for the simulated detections (i.e. complete galaxy catalogs and perfect redshift measurements). As expected, GW170817 provides an unusually good statistical H_0 constraint.

For the purposes of this calculation, we use the K-band luminosity-weighted $L_K > 0.1L_K^\star$ posterior shown in the right panel of Figure 7.6 as a representative posterior for the statistical GW170817 H_0 measurement. Over the wider prior $H_0 \in [10, 220]$ km s^{-1} Mpc^{-1} shown, the information difference between the posterior and the prior is 0.67 bits. The counterpart GW170817 H_0 measurement (dot-dashed pink curve in Figure 7.5) has an information gain of 1.55 bits with respect to the wide prior.

7.7 Conclusion

We perform a statistical standard siren measurement of the Hubble constant with GW170817. This analysis is the first application of the measurement originally proposed over 30 years ago by [219]. We find that the excellent localization of GW170817, together with the large scale structure causing galaxies to cluster into distinct groups, enables an informative measurement of H_0 even in the absence of a unique host galaxy identification. Including generic and flexible assumptions regarding the luminosities of BNS host galaxies, we find a peak at $H_0 \approx 76$ km s^{-1} Mpc^{-1} at ~ 2.4–3.7 times the prior probability density. We find the possibility of improved constraints when weighting the potential host galaxies by stellar mass and star-formation rate. Including all galaxies brighter than 0.01 $L_B^\star$ (including 99% of the total blue luminosity) we find $H_0 = 76^{+48}_{-23}$ km s^{-1} Mpc^{-1}, or $H_0 = 76^{+45}_{-21}$ km s^{-1} Mpc^{-1} when applying B-band luminosity weights (a proxy for star-formation rate). Weighting all

galaxies brighter than 0.005 $L_K^\star$ by their K-band luminosity (a proxy for stellar mass), we find $H_0 = 76^{+41}_{-19}$ km s^{-1} Mpc^{-1}.

Although this statistical standard siren measurement of H_0 is less precise than the counterpart measurement of $H_0 = 76^{+19}_{-13}$ km s^{-1} Mpc^{-1} (for a flat prior and utilizing the distance ansatz in the 3-dimensional sky map; see §7.6), it nonetheless shows that interesting constraints on cosmological parameters are possible from gravitational-wave sources even in the absence of an optical counterpart and an identification of the unique host galaxy [219, 81, 70, 124]. Although detailed studies find that the measurement of cosmological parameters from the counterpart approach is likely to surpass the statistical approach [70], the statistical approach offers an important cross-validation of the counterpart standard siren measurements. Furthermore, the statistical approach holds particular promise for binary black hole sources, which are detected at higher rates than binary neutron star systems and are expected to lack electromagnetic counterparts. The inferior quality of the individual H_0 measurements for binary black holes (because of the larger localization volumes) may be compensated for by the improved quantity due to the higher detection rates. The binary black holes will also be detected at much greater distances, and in addition to measuring H_0 may constrain additional cosmological parameters such as the equation of state of the dark energy.

7.8 Additional Analysis Details

7.8.1 Statistical H_0 likelihood

In this appendix we derive the H_0 posterior probability distribution function from Equation 7.1. We write the likelihood for GW and EM data, $x_{\rm GW}$ and $x_{\rm EM}$, given a value of H_0 as:

$$p(x_{\rm GW}, x_{\rm EM}|H_0) = \frac{\int p(x_{\rm GW}, x_{\rm EM}, D_L, \Omega, z|H_0)\, dD_L\, d\Omega\, dz}{\beta(H_0)}, \tag{7.9}$$

and factor the numerator as:

$$\begin{aligned}
&\int p(x_{\rm GW}, x_{\rm EM}, D_L, \Omega, z|H_0)\, dD_L\, d\Omega\, dz \\
&= \int p(x_{\rm GW}|D_L, \Omega) p(x_{\rm EM}|z, \Omega) p(D_L|z, H_0) p_0(z, \Omega)\, dD_L\, d\Omega\, dz \\
&= \int p(x_{\rm GW}|D_L, \Omega) p(x_{\rm EM}|z, \Omega) \delta(D_L - \hat{D}_L(z, H_0)) p_0(z, \Omega)\, dD_L\, d\Omega\, dz \\
&= \int p(x_{\rm GW}|\hat{D}_L(z, H_0), \Omega) p(x_{\rm EM}|z, \Omega) p_0(z, \Omega)\, d\Omega\, dz.
\end{aligned} \tag{7.10}$$

The H_0 posterior is related to the likelihood in Equation 7.10 by a prior:

$$p(H_0 \mid x_{\rm GW}, x_{\rm EM}) = p_0(H_0) p(x_{\rm GW}, x_{\rm EM}|H_0). \tag{7.11}$$

This equation is identical to Equation 7.1 in the main text. The normalization term $\beta(H_0)$ is given by [see 179, 70]:

$$\begin{aligned}
\beta(H_0) &= \int_{x_{\rm GW}>{\rm thresh}} \int_{x_{\rm EM}>{\rm thresh}} \int p(x_{\rm GW}, x_{\rm EM}, D_L, \Omega, z|H_0)\, dD_L\, dz\, d\Omega\, dx_{\rm GW}\, dx_{\rm EM} \\
&= \int P_{\rm det}^{\rm GW}(\hat{D}_L(z, H_0), \Omega, z) P_{\rm det}^{\rm EM}(z, \Omega) p_0(z, \Omega)\, d\Omega\, dz \\
&= \int_0^{z_{\rm h}} \int\int P_{\rm det}^{\rm GW}(\hat{D}_L(z, H_0), \Omega, z) p_0(z, \Omega)\, d\Omega\, dz,
\end{aligned} \tag{7.12}$$

where we assume that only data that is above some threshold is detected, and we define:

$$P_{\rm det}^{\rm GW}(D_L, \Omega, z) \equiv \int_{d_{\rm GW}>{\rm thresh}} p(d_{\rm GW}|D_L, \Omega, z)\, dd_{\rm GW}, \tag{7.13}$$

and similarly:

$$\begin{aligned}
P_{\rm det}^{\rm EM}(z, \Omega) &\equiv \int_{d_{\rm EM}>{\rm thresh}} p(d_{\rm EM}|z, \Omega) dd_{\rm EM} \\
&= \mathcal{H}(z_{\rm h} - z),
\end{aligned} \tag{7.14}$$

where $\mathcal{H}$ is the Heaviside step function. We assume that the EM likelihood is constant with redshift up to a maximum (horizon) redshift, beyond which we assume there are no detectable host galaxies. In the statistical analysis in which the EM likelihood is assumed to be uninformative, z_h is equivalent to the maximum redshift of our galaxy catalog, or $z_{\rm max}$ defined before Equation 7.5. We calculate $P_{\rm det}^{\rm GW}$ by assuming a network signal-to-noise ratio threshold of 12 for detection, a monochromatic BNS mass distribution of 1.4–1.4 $M_\odot$, zero spins, and isotropic inclination angles.

In practice, for nearby BNS sources, the term $\beta(H_0)$ is insensitive to the precise details of this calculation or to the choice of $z_{\rm h} \gtrsim 0.2$, and is essentially $\beta(H_0) \sim H_0^3$. This can be seen as follows. In LIGO-Virgo's second observing run, detectable BNS sources were within $\sim 100\,{\rm Mpc}$ [7]. For H_0 values within our prior range, this corresponds to redshifts $z \lesssim 0.07$, which is much smaller than the maximum detectable galaxy redshift, and so we can work in the limit $z_{\rm h} \to \infty$. We furthermore assume that the large-scale distribution of galaxies is uniform in comoving volume and we use the low-redshift, linear approximation $H_0 = cz/\ D_L$. At the low redshifts of detected BNS events, the redshifting of the GW signal in the detectors is negligible, and so we assume that $P_{\rm det}^{\rm GW}$ depends only on D_L and Ω, and is independent of z. With these approximations, we apply a different chain rule factorization to Equation 7.10 and write:

$$p(x_{\rm GW}, dx_{\rm EM}|H_0)\alpha(H_0) = \int p(x_{GW}|D_L,\Omega)p_0(D_L,\Omega)p(x_{EM}|\hat{z}(D_L,H_0),\Omega)\,dD_L\,d\Omega, \tag{7.15}$$

where $\alpha(H_0)$ is a normalization term analogous to $\beta(H_0)$. With this factorization, we can follow the steps in Equation 7.12 to write $\alpha(H_0)$ as:

$$\alpha(H_0) = \int P_{\rm det}^{\rm GW}(D_L,\Omega)p_0(D_L,\Omega)P_{\rm det}^{\rm EM}(\hat{z}(D_L,H_0),\Omega)\,d\Omega\,dD_L, \tag{7.16}$$

but this is now a constant (independent of H_0) because $P_{\rm det}^{\rm EM}(z,\Omega) = 1$. We can then do a

change of variables $dD_L = c/H_0\, dz$, and if we assume that $p_0(D_L, \Omega) \propto D_L^2$, we get:

$$p(x_{\rm GW}, x_{\rm EM}|H_0) \propto \frac{1}{H_0^3} \int p(x_{\rm GW}|\hat{D}_L(z, H_0), \Omega) p(x_{\rm EM}|z, \Omega) p_0(z, \Omega)\, d\Omega\, dz. \tag{7.17}$$

(Here we have dropped $\alpha(H_0)$ because it is a constant.) This is equivalent to Equation 7.10 if we set $\beta(H_0) \propto H_0^3$.

7.8.2 GW170817-like events

In order to explore whether the large-scale structure in the GW170817 localization volume, and the resulting statistical H_0 posterior, is typical for galaxies at $z \sim 0.01$, we rotate the true GW170817 sky map to different galaxies in the the MICE simulated catalog and repeat the statistical H_0 measurement. We assume that unlike for real galaxies in the GW170817 localization volume, no detailed observations have been carried out to measure the peculiar velocity field and apply group corrections. We therefore use the un-corrected redshifts given in the MICE catalog, which include a peculiar velocity contribution. The distribution of peculiar velocities is approximately described by a Gaussian of width 400 km/s, and we incorporate this uncertainty in the simulated H_0 measurements. Figure 7.7 shows the results for 20 realizations of the GW170817 localization volume centered on different galaxies in the MICE catalog. We see that the true GW170817 statistical H_0 measurement (we once again use the K-band luminosity-weighted $L_K > 0.1L_K^\star$ posterior shown in the right panel of Figure 7.6 as a representative posterior) is fairly typical among the different realizations. Over 50 different realizations, the information, given by the difference in Shannon entropy between prior and posterior, is $0.43^{+0.43}_{-0.19}$ bits (median and symmetric 90% intervals), whereas the information for the true GW170817 measurement is 0.67 bits. If we lower the peculiar velocity uncertainty in the simulations from 400 km/s to 200 km/s, the GW1708170-like posteriors become slightly more informative on average, with a typical information gain of

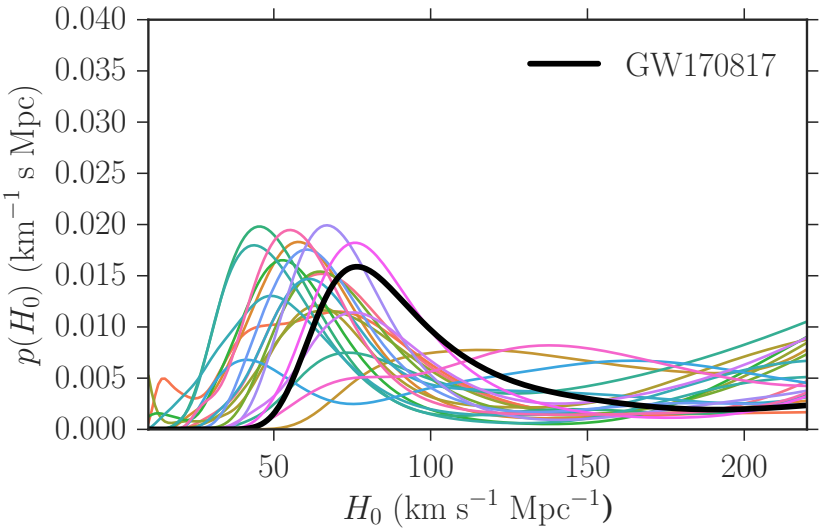

Figure 7.7: H_0 **posteriors for 20 realizations** of the GW170817 3-dimensional sky map centered on different galaxies in the MICE simulated galaxy catalog, assuming $H_0 = 70$ km s^{-1} Mpc^{-1}, and incorporating realistic (uncorrected) peculiar velocities with 1-σ uncertainties of 400 km/s. The real H_0 posterior using the GLADE galaxy catalog is shown in black. It is a typical result for a source with such a small localization volume, as such sources tend to produce a single major peak in the H_0 posterior.

$0.57^{+0.42}_{-0.27}$ bits.

REFERENCES

[1] J. Aasi et al. Advanced LIGO. *Classical and Quantum Gravity*, 32(7):074001, April 2015.

[2] B. P. Abbott et al. Astrophysical Implications of the Binary Black-hole Merger GW150914. *ApJL*, 818:L22, February 2016.

[3] B. P. Abbott et al. Binary Black Hole Mergers in the First Advanced LIGO Observing Run. *Physical Review X*, 6(4):041015, October 2016.

[4] B. P. Abbott et al. GW150914: First results from the search for binary black hole coalescence with Advanced LIGO. *Phys. Rev. D*, 93(12):122003, June 2016.

[5] B. P. Abbott et al. GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence. *Phys. Rev. Lett.*, 116(24):241103, Jun 2016.

[6] B. P. Abbott et al. Properties of the Binary Black Hole Merger GW150914. *Physical Review Letters*, 116(24):241102, June 2016.

[7] B. P. Abbott et al. Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO and Advanced Virgo. *Living Reviews in Relativity*, 19:1, February 2016.

[8] B. P. Abbott et al. The Rate of Binary Black Hole Mergers Inferred from Advanced LIGO Observations Surrounding GW150914. *ApJL*, 833:L1, December 2016.

[9] B. P. Abbott et al. A gravitational-wave standard siren measurement of the Hubble constant. *Nature*, 551:85–88, November 2017.

[10] B. P. Abbott et al. All-sky search for short gravitational-wave bursts in the first Advanced LIGO run. *Phys. Rev. D*, 95(4):042003, February 2017.

[11] B. P. Abbott et al. GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2. *Physical Review Letters*, 118(22):221101, June 2017.

[12] B. P. Abbott et al. GW170608: Observation of a 19 Solar-mass Binary Black Hole Coalescence. *ApJ*, 851(2):L35, Dec 2017.

[13] B. P. Abbott et al. GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence. *Physical Review Letters*, 119(14):141101, October 2017.

[14] B. P. Abbott et al. Gw170817: Observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett.*, 119:161101, Oct 2017.

[15] B. P. Abbott et al. Multi-messenger Observations of a Binary Neutron Star Merger. *ApJ*, 848:L12, October 2017.

[16] B. P. Abbott et al. Search for intermediate mass black hole binaries in the first observing run of Advanced LIGO. *ArXiv e-prints*, April 2017.

[17] B. P. Abbott et al. GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs. *arXiv e-prints*, page arXiv:1811.12907, Nov 2018.

[18] B. P. Abbott et al. Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Reviews in Relativity*, 21(1):3, Apr 2018.

[19] B. P. Abbott et al. Binary Black Hole Population Properties Inferred from the First and Second Observing Runs of Advanced LIGO and Advanced Virgo. *ApJ*, 882(2):L24, sep 2019.

[20] B. P. Abbott et al. Properties of the Binary Neutron Star Merger GW170817. *Physical Review X*, 9(1):011001, Jan 2019.

[21] B.P. Abbott et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116(6):061102, February 2016.

[22] B.P. Abbott et al. GW190412: Observation of a Binary-Black-Hole Coalescence with Asymmetric Masses. *arXiv e-prints*, page arXiv:2004.08342, April 2020.

[23] F. Acernese et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, 32(2):024001, Jan 2015.

[24] P. A. R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. *A&A*, 594:A13, September 2016.

[25] Hiroaki Aihara, Carlos Allende Prieto, Deokkeun An, Scott F. Anderson, Éric Aubourg, Eduardo Balbinot, Timothy C. Beers, Andreas A. Berlind, Steven J. Bickerton, Dmitry Bizyaev, Michael R. Blanton, John J. Bochanski, Adam S. Bolton, Jo Bovy, W. N. Brandt, J. Brinkmann, Peter J. Brown, Joel R. Brownstein, Nicolas G. Busca, Heather Campbell, Michael A. Carr, Yanmei Chen, Cristina Chiappini, Johan Comparat, Natalia Connolly, Marina Cortes, Rupert A. C. Croft, Antonio J. Cuesta, Luiz N. da Costa, James R. A. Davenport, Kyle Dawson, Saurav Dhital, Anne Ealet, Garrett L. Ebelke, Edward M. Edmondson, Daniel J. Eisenstein, Stephanie Escoffier, Massimiliano Esposito, Michael L. Evans, Xiaohui Fan, Bruno Femenía Castellá, Andreu Font- Ribera, Peter M. Frinchaboy, Jian Ge, Bruce A. Gillespie, G. Gilmore, Jonay I. González Hernández, J. Richard Gott, Andrew Gould, Eva K. Grebel, James E. Gunn, Jean-Christophe Hamilton, Paul Harding, David W. Harris, Suzanne L. Hawley, Frederick R. Hearty, Shirley Ho, David W. Hogg, Jon A. Holtzman, Klaus Honscheid, Naohisa Inada, Inese I. Ivans, Linhua Jiang, Jennifer A. Johnson, Cathy Jordan, Wendell P. Jordan, Eyal A. Kazin, David Kirkby, Mark A. Klaene, G. R. Knapp, Jean-Paul Kneib, C. S. Kochanek, Lars Koesterke, Juna A. Kollmeier,

Richard G. Kron, Hubert Lampeitl, Dustin Lang, Jean-Marc Le Goff, Young Sun Lee, Yen-Ting Lin, Daniel C. Long, Craig P. Loomis, Sara Lucatello, Britt Lundgren, Robert H. Lupton, Zhibo Ma, Nicholas MacDonald, Suvrath Mahadevan, Marcio A. G. Maia, Martin Makler, Elena Malanushenko, Viktor Malanushenko, Rachel Mandelbaum, Claudia Maraston, Daniel Margala, Karen L. Masters, Cameron K. McBride, Peregrine M. McGehee, Ian D. McGreer, Brice Ménard, Jordi Miralda-Escudé, Heather L. Morrison, F. Mullally, Demitri Muna, Jeffrey A. Munn, Hitoshi Murayama, Adam D. Myers, Tracy Naugle, Angelo Fausti Neto, Duy Cuong Nguyen, Robert C. Nichol, Robert W. O'Connell, Ricardo L. C. Ogando, Matthew D. Olmstead, Daniel J. Oravetz, Nikhil Padmanabhan, Nathalie Palanque- Delabrouille, Kaike Pan, Parul Pandey, Isabelle Pâris, Will J. Percival, Patrick Petitjean, Robert Pfaffenberger, Janine Pforr, Stefanie Phleps, Christophe Pichon, Matthew M. Pieri, Francisco Prada, Adrian M. Price-Whelan, M. Jordan Raddick, Beatriz H. F. Ramos, Céline Reylé, James Rich, Gordon T. Richards, Hans-Walter Rix, Annie C. Robin, Helio J. Rocha-Pinto, Constance M. Rockosi, Natalie A. Roe, Emmanuel Rollinde, Ashley J. Ross, Nicholas P. Ross, Bruno M. Rossetto, Ariel G. Sánchez, Conor Sayres, David J. Schlegel, Katharine J. Schlesinger, Sarah J. Schmidt, Donald P. Schneider, Erin Sheldon, Yiping Shu, Jennifer Simmerer, Audrey E. Simmons, Thirupathi Sivarani, Stephanie A. Snedden, Jennifer S. Sobeck, Matthias Steinmetz, Michael A. Strauss, Alexander S. Szalay, Masayuki Tanaka, Aniruddha R. Thakar, Daniel Thomas, Jeremy L. Tinker, Benjamin M. Tofflemire, Rita Tojeiro, Christy A. Tremonti, Jan Vandenberg, M. Vargas Magaña, Licia Verde, Nicole P. Vogt, David A. Wake, Ji Wang, Benjamin A. Weaver, David H. Weinberg, Martin White, Simon D. M. White, Brian Yanny, Naoki Yasuda, Christophe Yeche, and Idit Zehavi. The Eighth Data Release of the Sloan Digital Sky Survey: First Data from SDSS-III. *The Astrophysical Journal Supplement Series*, 193:29, April 2011.

[26] Bruce Allen, Warren G. Anderson, Patrick R. Brady, Duncan A. Brown, and Jolien D. E. Creighton. FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries. *Phys. Rev. D*, 85(12):122006, Jun 2012.

[27] P. Amaro-Seoane and X. Chen. Relativistic mergers of black hole binaries have large, similar masses, low spins and are circular. *MNRAS*, 458:3075–3082, May 2016.

[28] T. W. Anderson and D. A. Darling. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 23(2):193–212, June 1952.

[29] F. Antonini and F. A. Rasio. Merging Black Hole Binaries in Galactic Nuclei: Implications for Advanced-LIGO Detections. *ApJ*, 831:187, November 2016.

[30] M. Arca Sedda and M. Benacquista. Using final black hole spins and masses to infer the formation history of the observed population of gravitational wave sources. *MNRAS*, 482(3):2991–3010, Jan 2019.

[31] I. Arcavi, C. McCully, G. Hosseinzadeh, D. A. Howell, S. Vasylyev, D. Poznanski, M. Zaltzman, D. Maoz, L. Singer, S. Valenti, D. Kasen, J. Barnes, T. Piran, and W.-f. Fong. Optical Follow-up of Gravitational-wave Events with Las Cumbres Observatory. *ApJL*, 848:L33, October 2017.

[32] Abbas Askar, Magdalena Szkudlarek, Dorota Gondek-Rosińska, Mirek Giersz, and Tomasz Bulik. MOCCA-SURVEY Database - I. Coalescing binary black holes originating from globular clusters. *MNRAS*, 464(1):L36–L40, Jan 2017.

[33] Astropy Collaboration, T. P. Robitaille, E. J. Tollerud, P. Greenfield, M. Droettboom, E. Bray, T. Aldcroft, M. Davis, A. Ginsburg, A. M. Price-Whelan, W. E. Kerzendorf, A. Conley, N. Crighton, K. Barbary, D. Muna, H. Ferguson, F. Grollier, M. M. Parikh, P. H. Nair, H. M. Unther, C. Deil, J. Woillez, S. Conseil, R. Kramer, J. E. H. Turner, L. Singer, R. Fox, B. A. Weaver, V. Zabalza, Z. I. Edwards, K. Azalee Bostroem, D. J. Burke, A. R. Casey, S. M. Crawford, N. Dencheva, J. Ely, T. Jenness, K. Labrie, P. L. Lim, F. Pierfederici, A. Pontzen, A. Ptak, B. Refsdal, M. Servillat, and O. Streicher. Astropy: A community Python package for astronomy. *aap*, 558:A33, October 2013.

[34] Yang Bai, Vernon Barger, and Sida Lu. Measuring the Black Hole Mass Spectrum from Redshifts of aLIGO Binary Merger Events. *arXiv e-prints*, page arXiv:1802.04909, Feb 2018.

[35] Jim W. Barrett, Sebastian M. Gaebel, Coenraad J. Neijssel, Alejandro Vigna-Gómez, Simon Stevenson, Christopher P. L. Berry, Will M. Farr, and Ilya Mandel. Accuracy of inference on the physics of binary evolution from gravitational-wave observations. *MNRAS*, 477(4):4685–4695, Jul 2018.

[36] K. Belczynski, T. Bulik, and B. Rudak. The First Stellar Binary Black Holes: The Strongest Gravitational Wave Burst Sources. *ApJL*, 608:L45–L48, June 2004.

[37] K. Belczynski, A. Buonanno, M. Cantiello, C. L. Fryer, D. E. Holz, I. Mandel, M. C. Miller, and M. Walczak. The Formation and Gravitational-wave Detection of Massive Stellar Black Hole Binaries. *ApJ*, 789:120, July 2014.

[38] K. Belczynski, A. Heger, W. Gladysz, A. J. Ruiter, S. Woosley, G. Wiktorowicz, H.-Y. Chen, T. Bulik, R. O'Shaughnessy, D. E. Holz, C. L. Fryer, and E. Berti. The effect of pair-instability mass loss on black-hole mergers. *A&A*, 594:A97, October 2016.

[39] K. Belczynski, D. E. Holz, T. Bulik, and R. O'Shaughnessy. The first gravitational-wave source from the isolated evolution of two stars in the 40-100 solar mass range. *Nature*, 534:512–515, June 2016.

[40] K. Belczynski, J. Klencki, C. E. Fields, A. Olejak, E. Berti, G. Meynet, C. L. Fryer, D. E. Holz, R. O'Shaughnessy, D. A. Brown, T. Bulik, S. C. Leung, K. Nomoto, P. Madau, R. Hirschi, S. Jones, S. Mondal, M. Chruslinska, P. Drozda, D. Gerosa, Z. Doctor, M. Giersz, S. Ekstrom, C. Georgy, A. Askar, D. Wysocki, T. Natan, W. M.

Farr, G. Wiktorowicz, M. Coleman Miller, B. Farr, and J. P. Lasota. The evolutionary roads leading to low effective spins, high black hole masses, and O1/O2 rates of LIGO/Virgo binary black holes. *arXiv e-prints*, page arXiv:1706.07053, Jun 2017.

[41] K. Belczynski, S. Repetto, D. E. Holz, R. O'Shaughnessy, T. Bulik, E. Berti, C. Fryer, and M. Dominik. Compact Binary Merger Rates: Comparison with LIGO/Virgo Upper Limits. *ApJ*, 819:108, March 2016.

[42] K. Belczynski, G. Wiktorowicz, C. L. Fryer, D. E. Holz, and V. Kalogera. Missing Black Holes Unveil the Supernova Explosion Mechanism. *ApJ*, 757:91, September 2012.

[43] Krzysztof Belczynski, Michal Dominik, Tomasz Bulik, Richard O'Shaughnessy, Chris Fryer, and Daniel E. Holz. The effect of metallicity on the detection prospects for gravitational waves. *The Astrophysical Journal Letters*, 715(2):L138, 2010.

[44] Krzysztof Belczynski, Daniel E. Holz, Tomasz Bulik, and Richard O'Shaughnessy. The first gravitational-wave source from the isolated evolution of two stars in the 40-100 solar mass range. *Nature*, 534:512–515, June 2016.

[45] Krzysztof Belczynski, Serena Repetto, Daniel E. Holz, Richard O'Shaughnessy, Tomasz Bulik, Emanuele Berti, Christopher Fryer, and Michal Dominik. Compact Binary Merger Rates: Comparison with LIGO/Virgo Upper Limits. *ApJ*, 819(2):108, Mar 2016.

[46] E. F. Bell, D. H. McIntosh, N. Katz, and M. D. Weinberg. The Optical and Near-Infrared Properties of Galaxies. I. Luminosity and Stellar Mass Functions. *apjs*, 149:289–312, December 2003.

[47] E. Berti and M. Volonteri. Cosmological Black Hole Spin Evolution by Mergers and Accretion. *ApJ*, 684:822–828, September 2008.

[48] Simeon Bird, Ilias Cholis, Julian B. Muñoz, Yacine Ali-Haïmoud, Marc Kamionkowski, Ely D. Kovetz, Alvise Raccanelli, and Adam G. Riess. Did LIGO Detect Dark Matter? *Phys. Rev. Lett.*, 116(20):201301, May 2016.

[49] J. R. Bond, W. D. Arnett, and B. J. Carr. The evolution and fate of Very Massive Objects. *ApJ*, 280:825–847, May 1984.

[50] Yann Bouffanais, Michela Mapelli, Davide Gerosa, Ugo N. Di Carlo, Nicola Giacobbo, Emanuele Berti, and Vishal Baibhav. Constraining the fraction of binary black holes formed in isolation and young star clusters with gravitational-wave data. *arXiv e-prints*, page arXiv:1905.11054, May 2019.

[51] Tom Broadhurst, Jose M. Diego, and III Smoot, George. Reinterpreting Low Frequency LIGO/Virgo Events as Magnified Stellar-Mass Black Holes at Cosmological Distances. *arXiv e-prints*, page arXiv:1802.05273, Feb 2018.

[52] A. Buonanno, L. E. Kidder, and L. Lehner. Estimating the final spin of a binary black hole coalescence. *Phys. Rev. D*, 77(2):026004, January 2008.

[53] Erminia Calabrese, Nicholas Battaglia, and David N. Spergel. Testing gravity with gravitational wave source counts. *cqg*, 33(16):165004, August 2016.

[54] J. Calderón Bustillo, F. Salemi, T. dal Canton, and K. Jani. Sensitivity of gravitational wave searches to the full signal of intermediate mass black hole binaries during the LIGO O1 Science Run. *ArXiv e-prints*, November 2017.

[55] Juan Calderón Bustillo, Pablo Laguna, and Deirdre Shoemaker. Detectability of gravitational waves from binary black holes: Impact of precession and higher modes. *Phys. Rev. D*, 95:104038, May 2017.

[56] Thomas Callister, Maya Fishbach, Daniel Holz, and Will Farr. Shouts and Murmurs: Combining Individual Gravitational-Wave Sources with the Stochastic Background to Measure the History of Binary Black Hole Mergers. *arXiv e-prints*, page arXiv:2003.12152, March 2020.

[57] K. Cannon, R. Cariou, A. Chapman, M. Crispin-Ortuzar, N. Fotopoulos, M. Frei, C. Hanna, E. Kara, D. Keppel, L. Liao, S. Privitera, A. Searle, L. Singer, and A. Weinstein. Toward Early-warning Detection of Gravitational Waves from Compact Binary Coalescence. *ApJ*, 748:136, April 2012.

[58] Zhoujian Cao, Li-Fang Li, and Yan Wang. Gravitational lensing effects on parameter estimation in gravitational wave detection with advanced detectors. *Phys. Rev. D*, 90(6):062003, Sep 2014.

[59] C. Capano, Y. Pan, and A. Buonanno. Impact of higher harmonics in searching for gravitational waves from nonspinning binary black holes. *Phys. Rev. D*, 89(10):102003, May 2014.

[60] Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 2017.

[61] J. Carretero, F. J. Castander, E. Gaztañaga, M. Crocce, and P. Fosalba. An algorithm to build mock galaxy catalogues using MICE simulations. *MNRAS*, 447:646–670, February 2015.

[62] Jorge Carretero et al. CosmoHub and SciPIC: Massive cosmological data analysis, distribution and generation using a Big Data platform. *PoS*, EPS-HEP2017:488, 2017.

[63] J. Carrick, S. J. Turnbull, G. Lavaux, and M. J. Hudson. Cosmological parameters from the comparison of peculiar velocities with predictions from the 2M++ density field. *MNRAS*, 450:317–332, June 2015.

[64] Deepto Chakrabarty, Edward H. Morgan, Michael P. Muno, Duncan K. Galloway, Rudy Wijnands, Michiel van der Klis, and Craig B. Markwardt. Nuclear-powered millisecond pulsars and the maximum spin frequency of neutron stars. *Nature*, 424(6944):42–44, Jul 2003.

[65] Sourav Chatterjee, Carl L. Rodriguez, Vicky Kalogera, and Frederic A. Rasio. Dynamical Formation of Low-mass Merging Black Hole Binaries like GW151226. *ApJ*, 836(2):L26, Feb 2017.

[66] Katerina Chatziioannou, Roberto Cotesta, Sudarshan Ghonge, Jacob Lange, Ken K. Y. Ng, Juan Calderon Bustillo, James Clark, Carl-Johan Haster, Sebastian Khan, Michael Puerrer, Vivien Raymond, Salvatore Vitale, Nousha Afshari, Stanislav Babak, Kevin Barkett, Jonathan Blackman, Alejandro Bohe, Michael Boyle, Alessandra Buonanno, Manuela Campanelli, Gregorio Carullo, Tony Chu, Eric Flynn, Heather Fong, Alyssa Garcia, Matthew Giesler, Maria Haney, Mark Hannam, Ian Harry, James Healy, Daniel Hemberger, Ian Hinder, Karan Jani, Bhavesh Khamersa, Lawrence E. Kidder, Prayush Kumar, Pablo Laguna, Carlos O. Lousto, Geoffrey Lovelace, Tyson B. Littenberg, Lionel London, Margaret Millhouse, Laura K. Nuttall, Frank Ohme, Richard O'Shaughnessy, Serguei Ossokine, Francesco Pannarale, Patricia Schmidt, Harald P. Pfeiffer, Mark A. Scheel, Lijing Shao, Deirdre Shoemaker, Bela Szilagyi, Andrea Taracchini, Saul A. Teukolsky, and Yosef Zlochower. On the properties of the massive binary black hole merger GW170729. *arXiv e-prints*, page arXiv:1903.06742, Mar 2019.

[67] H.-Y. Chen and D. E. Holz. The Loudest Gravitational Wave Events. *ArXiv e-prints*, September 2014.

[68] H.-Y. Chen and D. E. Holz. Finding the One: Identifying the Host Galaxies of Gravitational-Wave Sources. *ArXiv e-prints*, December 2016.

[69] H.-Y. Chen and D. E. Holz. Facilitating Follow-up of LIGO-Virgo Events Using Rapid Sky Localization. *ApJ*, 840:88, May 2017.

[70] Hsin-Yu Chen, Maya Fishbach, and Daniel E. Holz. A two per cent Hubble constant measurement from standard sirens within five years. *Nature*, 562:545–547, October 2018.

[71] Hsin-Yu Chen, Daniel E. Holz, John Miller, Matthew Evans, Salvatore Vitale, and Jolien Creighton. Distance measures in gravitational-wave astrophysics and cosmology. *arXiv e-prints*, page arXiv:1709.08079, Sep 2017.

[72] Martyna Chruslinska, Krzysztof Belczynski, Jakub Klencki, and Matthew Benacquista. Double neutron stars: merger rates revisited. *MNRAS*, 474(3):2937–2958, Mar 2018.

[73] S. Clesse and J. García-Bellido. The clustering of massive Primordial Black Holes as Dark Matter: Measuring their mass distribution with advanced LIGO. *Physics of the Dark Universe*, 15:142–147, March 2017.

[74] D. A. Coulter, R. J. Foley, C. D. Kilpatrick, M. R. Drout, A. L. Piro, B. J. Shappee, M. R. Siebert, J. D. Simon, N. Ulloa, D. Kasen, B. F. Madore, A. Murguia-Berthier, Y. C. Pan, J. X. Prochaska, E. Ramirez-Ruiz, A. Rest, and C. Rojas-Bravo. Swope Supernova Survey 2017a (SSS17a), the optical counterpart to a gravitational wave source. *Science*, 358:1556–1558, December 2017.

[75] M. Crocce, F. J. Castander, E. Gaztañaga, P. Fosalba, and J. Carretero. The MICE Grand Challenge lightcone simulation - II. Halo and galaxy catalogues. *MNRAS*, 453:1513–1530, October 2015.

[76] Liang Dai, Tejaswi Venumadhav, and Kris Sigurdson. Effect of lensing magnification on the apparent distribution of black hole mergers. *Phys. Rev. D*, 95(4):044011, Feb 2017.

[77] T. Dal Canton and I. W. Harry. Designing a template bank to observe compact binary coalescences in Advanced LIGO's second observing run. *ArXiv e-prints*, May 2017.

[78] G. Dálya, G. Galgóczi, L. Dobos, Z. Frei, I. S. Heng, R. Macas, C. Messenger, P. Raffai, and R. S. de Souza. GLADE: A Galaxy Catalogue for Multi-Messenger Searches in the Advanced Gravitational-Wave Detector Era. *ArXiv e-prints*, April 2018.

[79] S. E. de Mink and I. Mandel. The chemically homogeneous evolutionary channel for binary black hole mergers: rates and properties of gravitational-wave events detectable by advanced LIGO. *MNRAS*, 460:3545–3553, August 2016.

[80] Cédric Deffayet and Kristen Menou. Probing Gravity with Spacetime Sirens. *ApJL*, 668(2):L143–L146, October 2007.

[81] W. Del Pozzo. Inference of cosmological parameters from gravitational waves: Applications to second generation interferometers. *Phys. Rev. D*, 86(4):043011, August 2012.

[82] W. Del Pozzo, C. P. L. Berry, A. Ghosh, T. S. F. Haines, L. P. Singer, and A. Vecchio. Dirichlet Process Gaussian-mixture model: An application to localizing coalescing binary neutron stars with gravitational-wave observations. *MNRAS*, June 2018.

[83] Ugo N. Di Carlo, Nicola Giacobbo, Michela Mapelli, Mario Pasquato, Mario Spera, Long Wang, and Francesco Haardt. Merging black holes in young star clusters. *arXiv e-prints*, page arXiv:1901.00863, Jan 2019.

[84] James M. Dickey. The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters. *Ann. Math. Statist.*, 42:204–223, 1971.

[85] Zoheyr Doctor, Daniel Wysocki, Richard O'Shaughnessy, Daniel E. Holz, and Ben Farr. Black Hole Coagulation: Modeling Hierarchical Mergers in Black Hole Populations. *arXiv e-prints*, page arXiv:1911.04424, Nov 2019.

[86] M. Dominik, K. Belczynski, C. Fryer, D. E. Holz, E. Berti, T. Bulik, I. Mandel, and R. O'Shaughnessy. Double Compact Objects. I. The Significance of the Common Envelope on Merger Rates. *ApJ*, 759:52, November 2012.

[87] M. Dominik, E. Berti, R. O'Shaughnessy, I. Mandel, K. Belczynski, C. Fryer, D. E. Holz, T. Bulik, and F. Pannarale. Double Compact Objects III: Gravitational-wave Detection Rates. *ApJ*, 806:263, June 2015.

[88] Michal Dominik, Krzysztof Belczynski, Christopher Fryer, Daniel E. Holz, Emanuele Berti, Tomasz Bulik, Ilya Mand el, and Richard O'Shaughnessy. Double Compact Objects. I. The Significance of the Common Envelope on Merger Rates. *ApJ*, 759(1):52, Nov 2012.

[89] Michal Dominik, Krzysztof Belczynski, Christopher Fryer, Daniel E. Holz, Emanuele Berti, Tomasz Bulik, Ilya Mand el, and Richard O'Shaughnessy. Double Compact Objects. II. Cosmological Merger Rates. *ApJ*, 779(1):72, Dec 2013.

[90] Michal Dominik, Krzysztof Belczynski, Christopher Fryer, Daniel E. Holz, Emanuele Berti, Tomasz Bulik, Ilya Mandel, and Richard O'Shaughnessy. Double Compact Objects. I. The Significance of the Common Envelope on Merger Rates. *ApJ*, 759(1):52, Nov 2012.

[91] Michal Dominik, Emanuele Berti, Richard O'Shaughnessy, Ilya Mandel, Krzysztof Belczynski, Christopher Fryer, Daniel E. Holz, Tomasz Bulik, and Francesco Pannarale. Double Compact Objects III: Gravitational-wave Detection Rates. *ApJ*, 806(2):263, Jun 2015.

[92] G. Dvali, G. Gabadadze, and M. Porrati. 4D gravity on a brane in 5D Minkowski space. *Physics Letters B*, 485:208–214, July 2000.

[93] J. J. Eldridge and E. R. Stanway. BPASS predictions for binary black hole mergers. *MNRAS*, 462:3302–3313, November 2016.

[94] J. J. Eldridge and E. R. Stanway. BPASS predictions for binary black hole mergers. *MNRAS*, 462(3):3302–3313, Nov 2016.

[95] Ben Farr, Daniel E. Holz, and Will M. Farr. Using Spin to Understand the Formation of LIGO and Virgo's Black Holes. *ApJ*, 854(1):L9, Feb 2018.

[96] W. M. Farr, I. Mandel, C. Aldridge, and K. Stroud. The Occurrence of Earth-Like Planets Around Other Stars. *ArXiv e-prints*, December 2014.

[97] W. M. Farr, N. Sravan, A. Cantrell, L. Kreidberg, C. D. Bailyn, I. Mandel, and V. Kalogera. The Mass Distribution of Stellar-mass Black Holes. *ApJ*, 741:103, November 2011.

[98] Will M. Farr, Maya Fishbach, Jiani Ye, and Daniel E. Holz. A Future Percent-level Measurement of the Hubble Expansion at Redshift 0.8 with Advanced LIGO. *ApJL*, 883(2):L42, Oct 2019.

[99] Will M. Farr, Simon Stevenson, M. Coleman Miller, Ilya Mand el, Ben Farr, and Alberto Vecchio. Distinguishing spin-aligned and isotropic black hole populations with gravitational waves. *Nature*, 548(7667):426–429, Aug 2017.

[100] Will M. Farr, Simon Stevenson, M. Coleman Miller, Ilya Mand el, Ben Farr, and Alberto Vecchio. Distinguishing spin-aligned and isotropic black hole populations with gravitational waves. *Nature*, 548(7667):426–429, Aug 2017.

[101] L. S. Finn and D. F. Chernoff. Observing binary inspiral in gravitational radiation: One interferometer. *Phys. Rev. D*, 47:2198–2219, March 1993.

[102] Lee Samuel Finn. Binary inspiral, gravitational radiation, and cosmology. *Phys. Rev. D*, 53(6):2878–2894, Mar 1996.

[103] Lee Samuel Finn and David F. Chernoff. Observing binary inspiral in gravitational radiation: One interferometer. *Phys. Rev. D*, 47(6):2198–2219, Mar 1993.

[104] M. Fishbach and etl al. A Standard Siren Measurement of the Hubble Constant from GW170817 without the Electromagnetic Counterpart. *ApJL*, 871(1):L13, January 2019.

[105] M. Fishbach and D. E. Holz. Where Are LIGO's Big Black Holes? *ApJL*, 851:L25, December 2017.

[106] M. Fishbach, D. E. Holz, and B. Farr. Are LIGO's Black Holes Made from Smaller Black Holes? *ApJL*, 840:L24, May 2017.

[107] Maya Fishbach, Will M. Farr, and Daniel E. Holz. The Most Massive Binary Black Hole Detections and the Identification of Population Outliers. *ApJL*, 891(2):L31, March 2020.

[108] Maya Fishbach and Daniel E. Holz. Picky Partners: The Pairing of Component Masses in Binary Black Hole Mergers. *ApJL*, 891(1):L27, March 2020.

[109] Maya Fishbach, Daniel E. Holz, and Will M. Farr. Does the Black Hole Merger Rate Evolve with Redshift? *ApJL*, 863(2):L41, Aug 2018.

[110] W. Fong and E. Berger. The locations of short gamma-ray bursts as evidence for compact object binary progenitors. *The Astrophysical Journal*, 776(1):18, 2013.

[111] Daniel Foreman-Mackey. corner.py: Scatterplot matrices in python. *The Journal of Open Source Software*, 24, 2016.

[112] Daniel Foreman-Mackey, David W. Hogg, and Timothy D. Morton. Exoplanet Population Inference and the Abundance of Earth Analogs from Noisy, Incomplete Catalogs. *ApJ*, 795(1):64, Nov 2014.

[113] P. Fosalba, M. Crocce, E. Gaztañaga, and F. J. Castander. The MICE grand challenge lightcone simulation - I. Dark matter clustering. *MNRAS*, 448:2987–3000, April 2015.

[114] P. Fosalba, E. Gaztañaga, F. J. Castander, and M. Crocce. The MICE Grand Challenge light-cone simulation - III. Galaxy lensing mocks from all-sky lensing maps. *MNRAS*, 447:1319–1332, February 2015.

[115] W. A. Fowler and F. Hoyle. Neutrino Processes and Pair Formation in Massive Stars and Supernovae. *apjs*, 9:201, December 1964.

[116] Jim Fuller and Linhao Ma. Most Black Holes Are Born Very Slowly Rotating. *ApJL*, 881(1):L1, Aug 2019.

[117] Juan García-Bellido. Massive Primordial Black Holes as Dark Matter and their detection with Gravitational Waves. In *Journal of Physics Conference Series*, volume 840 of *Journal of Physics Conference Series*, page 012032, May 2017.

[118] Juan García-Bellido, Savvas Nesseris, and Manuel Trashorras. Gravitational wave source counts at high redshift and in models with extra dimensions. 2016(7):021, July 2016.

[119] N. Gehrels, J. K. Cannizzo, J. Kanner, M. M. Kasliwal, S. Nissanke, and L. P. Singer. Galaxy Strategy for LIGO-Virgo Gravitational Wave Counterpart Searches. *ApJ*, 820:136, April 2016.

[120] D. Gerosa and E. Berti. Are merging black holes born from stellar collapse or previous mergers? *Phys. Rev. D*, 95(12):124046, June 2017.

[121] Archisman Ghosh, Walter Del Pozzo, and Parameswaran Ajith. Estimating parameters of binary black holes from gravitational-wave observations of their inspiral, merger, and ringdown. *Phys. Rev. D*, 94(10):104070, November 2016.

[122] Nicola Giacobbo, Michela Mapelli, and Mario Spera. Merging black hole binaries: the effects of progenitor's metallicity, mass-loss rate and Eddington factor. *MNRAS*, 474(3):2959–2974, Mar 2018.

[123] Gravitational Wave Open Science Center. Strain data release for gwtc-1: A gravitational-wave transient catalog of compact binary mergers observed by ligo and virgo during the first and second observing runs, 2018.

[124] Rachel Gray, Ignacio Magaña Hernandez, Hong Qi, Ankan Sur, Patrick R. Brady, Hsin-Yu Chen, Will M. Farr, Maya Fishbach, Jonathan R. Gair, Archisman Ghosh, Daniel E. Holz, Simone Mastrogiovanni, Christopher Messenger, Danièle A. Steer, and

John Veitch. Cosmological Inference using Gravitational Wave Standard Sirens: A Mock Data Challenge. *arXiv e-prints*, page arXiv:1908.06050, August 2019.

[125] O. A. Hannuksela, K. Haris, K. K. Y. Ng, S. Kumar, A. K. Mehta, D. Keitel, T. G. F. Li, and P. Ajith. Search for Gravitational Lensing Signatures in LIGO-Virgo Binary Black Hole Events. *ApJL*, 874(1):L2, Mar 2019.

[126] J. Healy, C. O. Lousto, and Y. Zlochower. Remnant mass, spin, and recoil from spin aligned black-hole binaries. *Phys. Rev. D*, 90(10):104004, November 2014.

[127] A. Heger and S. E. Woosley. The Nucleosynthetic Signature of Population III. *ApJ*, 567:532–543, March 2002.

[128] G. Hinshaw, J. L. Weiland, R. S. Hill, N. Odegard, D. Larson, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, N. Jarosik, E. Komatsu, M. R. Nolta, L. Page, D. N. Spergel, E. Wollack, M. Halpern, A. Kogut, M. Limon, S. S. Meyer, G. S. Tucker, and E. L. Wright. Five-Year Wilkinson Microwave Anisotropy Probe Observations: Data Processing, Sky Maps, and Basic Results. *apjs*, 180:225–245, February 2009.

[129] J. Hjorth, A. J. Levan, N. R. Tanvir, J. D. Lyman, R. Wojtak, S. L. Schrøder, I. Mandel, C. Gall, and S. H. Bruun. The Distance to NGC 4993: The Host Galaxy of the Gravitational-wave Event GW170817. *ApJL*, 848:L31, October 2017.

[130] K. Hoffmann, J. Bel, E. Gaztañaga, M. Crocce, P. Fosalba, and F. J. Castander. Measuring the growth of matter fluctuations with third-order galaxy correlations. *MNRAS*, 447:1724–1745, February 2015.

[131] F. Hofmann, E. Barausse, and L. Rezzolla. The Final Spin from Binary Black Holes in Quasi-circular Orbits. *ApJL*, 825:L19, July 2016.

[132] D. W. Hogg, A. D. Myers, and J. Bovy. Inferring the Eccentricity Distribution. *ApJ*, 725:2166–2175, December 2010.

[133] David W. Hogg. Distance measures in cosmology. *arXiv e-prints*, pages astro–ph/9905116, May 1999.

[134] David W. Hogg, Adam D. Myers, and Jo Bovy. Inferring the Eccentricity Distribution. *ApJ*, 725(2):2166–2175, Dec 2010.

[135] Daniel E. Holz and Scott A. Hughes. Using Gravitational-Wave Standard Sirens. *ApJ*, 629(1):15–22, August 2005.

[136] Daniel E. Holz and Eric V. Linder. Safety in Numbers: Gravitational Lensing Degradation of the Luminosity Distance-Redshift Relation. *ApJ*, 631(2):678–688, October 2005.

[137] Daniel E. Holz and Robert M. Wald. New method for determining cumulative gravitational lensing effects in inhomogeneous universes. *Phys. Rev. D*, 58(6):063501, September 1998.

[138] S. A. Hughes and R. D. Blandford. Black Hole Mass and Spin Coevolution by Mergers. *ApJL*, 585:L101–L104, March 2003.

[139] Jarrod R. Hurley, Anna C. Sippel, Christopher A. Tout, and Sverre J. Aarseth. A Dynamical Gravitational Wave Source in a Dense Cluster. *pasa*, 33:e036, Aug 2016.

[140] Sascha Husa, Sebastian Khan, Mark Hannam, Michael Pürrer, Frank Ohme, Xisco Jiménez Forteza, and Alejandro Bohé. Frequency-domain gravitational waves from nonprecessing black-hole binaries. i. new numerical waveforms and anatomy of the signal. *Phys. Rev. D*, 93:044006, Feb 2016.

[141] K. Inayoshi, R. Hirai, T. Kinugawa, and K. Hotokezaka. Formation pathway of Population III coalescing binary black holes through stable mass transfer. *ArXiv e-prints*, January 2017.

[142] Xisco Jiménez-Forteza, David Keitel, Sascha Husa, Mark Hannam, Sebastian Khan, and Michael Prrer. Hierarchical data-driven approach to fitting numerical relativity data for nonprecessing binary black holes with an application to final spin and radiated energy. *Phys. Rev.*, D95(6):064024, 2017.

[143] Roger W. Johnson. Estimating the size of a population. *Wiley Online Library*, Jun 2007.

[144] S. Karki, D. Tuyenbayev, S. Kandhasamy, B. P. Abbott, T. D. Abbott, E. H. Anders, J. Berliner, J. Betzwieser, C. Cahillane, L. Canete, C. Conley, H. P. Daveloza, N. De Lillo, J. R. Gleason, E. Goetz, K. Izumi, J. S. Kissel, G. Mendell, V. Quetschke, M. Rodruck, S. Sachdev, T. Sadecki, P. B. Schwinberg, A. Sottile, M. Wade, A. J. Weinstein, M. West, and R. L. Savage. The Advanced LIGO photon calibrators. *Review of Scientific Instruments*, 87:114503, November 2016.

[145] Robert E. Kass and Adrian E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

[146] M. Kesden. Can binary mergers produce maximally spinning black holes? *Phys. Rev. D*, 78(8):084030, October 2008.

[147] M. Kesden, U. Sperhake, and E. Berti. Final spins from the merger of precessing binary black holes. *Phys. Rev. D*, 81(8):084054, April 2010.

[148] Sebastian Khan, Sascha Husa, Mark Hannam, Frank Ohme, Michael Pürrer, Xisco Jiménez Forteza, and Alejandro Bohé. Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era. *Phys. Rev. D*, 93(4):044007, Feb 2016.

[149] Chase Kimball, Christopher P L Berry, and Vicky Kalogera. What GW170729's exceptional mass and spin tells us about its family tree. *arXiv e-prints*, page arXiv:1903.07813, Mar 2019.

[150] T. Kinugawa, K. Inayoshi, K. Hotokezaka, D. Nakauchi, and T. Nakamura. Possible indirect confirmation of the existence of Pop III massive stars by gravitational wave. *MNRAS*, 442:2963–2992, August 2014.

[151] J. Kissel, J. Betzwieser, X. Siemens, R. Savage, K. Kawabe, M. Wade, B. OReilly, K. Izumi, S. Karki, D. Tuyenbayev, D. Martynov, S. Kandhasamy, M. Fays, , and C. Cahillane. Advanced ligo sensitivity plots. *Tech. Rep. LIGO-G1500623*, 2016.

[152] J. Klencki, M. Moe, W. Gladysz, M. Chruslinska, D. E. Holz, and K. Belczynski. Impact of inter-correlated initial binary parameters on double black hole and neutron star mergers. *A&A*, 619:A77, Nov 2018.

[153] S. Klimenko, I. Yakushin, A. Mercer, and G. Mitselmakher. A coherent method for detection of gravitational wave bursts. *Classical and Quantum Gravity*, 25(11):114029, June 2008.

[154] Bence Kocsis, Teruaki Suyama, Takahiro Tanaka, and Shuichiro Yokoyama. Hidden Universality in the Merger Rate Distribution in the Primordial Black Hole Scenario. *ApJ*, 854(1):41, Feb 2018.

[155] Andrey Kolmogorov. Sulla determinazione empirica di una legge di distribuzione. *Giornale dellIstituto Italiano degli Attuari*, 4:83–91, 1933.

[156] E. Kourkchi and R. B. Tully. Galaxy Groups Within 3500 km s^{-1}. *ApJ*, 843:16, July 2017.

[157] Savvas M. Koushiappas and Abraham Loeb. Maximum redshift of gravitational wave merger events. *Phys. Rev. Lett.*, 119:221104, Nov 2017.

[158] M. B. N. Kouwenhoven, A. G. A. Brown, S. P. Goodwin, S. F. Portegies Zwart, and L. Kaper. Exploring the consequences of pairing algorithms for binary stars. *A&A*, 493:979–1016, January 2009.

[159] E. D. Kovetz, I. Cholis, P. C. Breysse, and M. Kamionkowski. Black hole mass function from gravitational wave measurements. *Phys. Rev. D*, 95(10):103010, May 2017.

[160] L. Kreidberg, C. D. Bailyn, W. M. Farr, and V. Kalogera. Mass Measurements of Black Holes in X-Ray Transients: Is There a Mass Gap? *ApJ*, 757:36, September 2012.

[161] A. Krolak and Bernard F. Schutz. Coalescing binaries—Probe of the universe. *General Relativity and Gravitation*, 19(12):1163–1171, December 1987.

[162] P. Kroupa. On the variation of the initial mass function. *MNRAS*, 322:231–246, April 2001.

[163] P. Kroupa, C. Weidner, J. Pflamm-Altenburg, I. Thies, J. Dabringhausen, M. Marks, and T. Maschberger. *The Stellar and Sub-Stellar Initial Mass Function of Simple and Composite Populations*, page 115. 2013.

[164] Pavel Kroupa and Tereza Jerabkova. The Impact of Binaries on the Stellar Initial Mass Function. *arXiv e-prints*, page arXiv:1806.10605, June 2018.

[165] Matthias U. Kruckow, Thomas M. Tauris, Norbert Langer, Michael Kramer, and Robert G. Izzard. Progenitors of gravitational wave mergers: binary evolution with the stellar grid-based code COMBINE. *MNRAS*, 481(2):1908–1949, Dec 2018.

[166] N. Langer and C. A. Norman. On the Collapsar Model of Long Gamma-Ray Bursts:Constraints from Cosmic Metallicity Evolution. *ApJL*, 638(2):L63–L66, February 2006.

[167] C. N. Leibler and E. Berger. The Stellar Ages and Masses of Short Gamma-ray Burst Host Galaxies: Investigating the Progenitor Delay Time Distribution and the Role of Mass and Star Formation in the Short Gamma-ray Burst Rate. *ApJ*, 725:1202–1214, December 2010.

[168] T. B. Littenberg, B. Farr, S. Coughlin, V. Kalogera, and D. E. Holz. Neutron Stars versus Black Holes: Probing the Mass Gap with LIGO/Virgo. *ApJL*, 807:L24, July 2015.

[169] Thomas J. Loredo. Accounting for Source Uncertainties in Analyses of Astronomical Survey Data. In *American Institute of Physics Conference Series*, volume 735, pages 195–206, November 2004.

[170] C. O. Lousto, H. Nakano, Y. Zlochower, and M. Campanelli. Statistical studies of spinning black-hole binaries. *Phys. Rev. D*, 81(8):084023, April 2010.

[171] Y. Lu, X. Yang, F. Shi, H. J. Mo, D. Tweed, H. Wang, Y. Zhang, S. Li, and S. H. Lim. Galaxy Groups in the 2Mass Redshift Survey. *ApJ*, 832:39, November 2016.

[172] C. L. MacLeod and C. J. Hogan. Precision of Hubble constant derived using black hole binary absolute distances and statistical redshift information. *Phys. Rev. D*, 77(4):043512, February 2008.

[173] Piero Madau and Mark Dickinson. Cosmic Star-Formation History. *araa*, 52:415–486, Aug 2014.

[174] I. Mandel and S. E. de Mink. Merging binary black holes formed through chemically homogeneous evolution in short-period stellar binaries. *MNRAS*, 458:2634–2647, May 2016.

[175] I. Mandel, W. M. Farr, A. Colonna, S. Stevenson, P. Tiňo, and J. Veitch. Model-independent inference on compact-binary observations. *MNRAS*, 465:3254–3260, March 2017.

[176] Ilya Mandel. Parameter estimation on gravitational waves from multiple coalescing binaries. *Phys. Rev. D*, 81(8):084029, Apr 2010.

[177] Ilya Mandel and Alison Farmer. Merging stellar-mass binary black holes. *arXiv e-prints*, page arXiv:1806.05820, Jun 2018.

[178] Ilya Mandel, Will M. Farr, Andrea Colonna, Simon Stevenson, Peter Tiňo, and John Veitch. Model-independent inference on compact-binary observations. *MNRAS*, 465(3):3254–3260, Mar 2017.

[179] Ilya Mandel, Will M. Farr, and Jonathan R. Gair. Extracting distribution parameters from multiple uncertain observations with selection biases. *MNRAS*, 486(1):1086–1093, Jun 2019.

[180] K. S. Mandel, G. Narayan, and R. P. Kirshner. Type Ia Supernova Light Curve Inference: Hierarchical Models in the Optical and Near-infrared. *ApJ*, 731:120, April 2011.

[181] Vuk Mandic, Simeon Bird, and Ilias Cholis. Stochastic Gravitational-Wave Background due to Primordial Binary Black Hole Mergers. *Phys. Rev. Lett.*, 117(20):201102, November 2016.

[182] M. Mapelli, L. Zampieri, E. Ripamonti, and A. Bressan. Dynamics of stellar black holes in young star clusters with different metallicities - I. Implications for X-ray binaries. *mnras*, 429(3):2298–2314, March 2013.

[183] Michela Mapelli. Massive black hole binaries from runaway collisions: the impact of metallicity. *MNRAS*, 459(4):3432–3446, Jul 2016.

[184] Michela Mapelli, Nicola Giacobbo, Emanuele Ripamonti, and Mario Spera. The cosmic merger rate of stellar black hole binaries from the Illustris simulation. *MNRAS*, 472(2):2422–2435, Dec 2017.

[185] P. Marchant, N. Langer, P. Podsiadlowski, T. M. Tauris, and T. J. Moriya. A new route towards merging massive black holes. *A&A*, 588:A50, April 2016.

[186] Pablo Marchant, Norbert Langer, Philipp Podsiadlowski, Thomas M. Tauris, and Takashi J. Moriya. A new route towards merging massive black holes. *A&A*, 588:A50, Apr 2016.

[187] B. McKernan, K. E. S. Ford, J. Bellovary, N. W. C. Leigh, Z. Haiman, B. Kocsis, W. Lyra, M.-M. MacLow, B. Metzger, M. O'Dowd, S. Endlich, and D. J. Rosen. On stellar-mass black hole mergers in AGN disks detectable with LIGO. *ArXiv e-prints*, February 2017.

[188] D. Merritt, M. Milosavljević, M. Favata, S. A. Hughes, and D. E. Holz. Consequences of Gravitational Radiation Recoil. *ApJL*, 607:L9–L12, May 2004.

[189] C. Messick, K. Blackburn, P. Brady, P. Brockill, K. Cannon, R. Cariou, S. Caudill, S. J. Chamberlin, J. D. E. Creighton, R. Everett, C. Hanna, D. Keppel, R. N. Lang, T. G. F. Li, D. Meacher, A. Nielsen, C. Pankow, S. Privitera, H. Qi, S. Sachdev, L. Sadeghian, L. Singer, E. G. Thomas, L. Wade, M. Wade, A. Weinstein, and K. Wiesner. Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data. *Phys. Rev. D*, 95(4):042001, February 2017.

[190] Erez Michaely and Hagai B. Perets. Gravitational-wave sources from mergers of binary black-holes catalyzed by fly-bys interactions in the field. *arXiv e-prints*, page arXiv:1902.01864, Feb 2019.

[191] M. C. Miller and J. M. Miller. The masses and spins of neutron stars and stellar-mass black holes. *physrep*, 548:1–34, January 2015.

[192] Maxwell Moe and Rosanne Di Stefano. Mind Your Ps and Qs: The Interrelation between Period (P) and Mass-ratio (Q) Distributions of Binary Stars. *The Astrophysical Journal Supplement Series*, 230:15, June 2017.

[193] R. Nair, S. Bose, and T. Deep Saini. Measuring the Hubble constant: gravitational wave observations meet galaxy clustering. *ArXiv e-prints*, April 2018.

[194] Ken K. Y. Ng, Kaze W. K. Wong, Tom Broadhurst, and Tjonnie G. F. Li. Precise ligo lensing rate predictions for binary black holes. *Phys. Rev. D*, 97:023012, Jan 2018.

[195] A. H. Nitz, T. Dent, T. Dal Canton, S. Fairhurst, and D. A. Brown. Detecting binary compact-object mergers with gravitational waves: Understanding and Improving the sensitivity of the PyCBC search. *ArXiv e-prints*, May 2017.

[196] Alexander H. Nitz, Thomas Dent, Gareth S. Davies, Sumit Kumar, Collin D. Capano, Ian Harry, Simone Mazzon, Laura Nuttall, Andrew Lundgren, and Marton Tápai. 2-OGC: Open Gravitational-wave Catalog of binary mergers from analysis of public Advanced LIGO and Virgo data. *arXiv e-prints*, page arXiv:1910.05331, Oct 2019.

[197] Masamune Oguri. Effect of gravitational lensing on the distribution of gravitational waves from distant binary black hole mergers. *MNRAS*, 480(3):3842–3855, 08 2018.

[198] R. M. O'Leary, Y. Meiron, and B. Kocsis. Dynamical Formation Signatures of Black Hole Binaries in the First Detected Mergers by LIGO. *ApJL*, 824:L12, June 2016.

[199] Ryan M. O'Leary, Frederic A. Rasio, John M. Fregeau, Natalia Ivanova, and Richard O'Shaughnessy. Binary Mergers and Growth of Black Holes in Dense Star Clusters. *ApJ*, 637(2):937–951, Feb 2006.

[200] F. Özel, D. Psaltis, R. Narayan, and J. E. McClintock. The Black Hole Mass Distribution in the Galaxy. *ApJ*, 725:1918–1927, December 2010.

[201] Feryal Özel, Dimitrios Psaltis, Ramesh Narayan, and Antonio Santos Villarreal. On the Mass Distribution and Birth Masses of Neutron Stars. *ApJ*, 757(1):55, Sep 2012.

[202] Yi Pan, Alessandra Buonanno, Andrea Taracchini, Lawrence E. Kidder, Abdul H. Mroué, Harald P. Pfeiffer, Mark A. Scheel, and Béla Szilágyi. Inspiral-merger-ringdown waveforms of spinning, precessing black-hole binaries in the effective-one-body formalism. *Phys. Rev. D*, 89(8):084006, Apr 2014.

[203] Kris Pardo, Maya Fishbach, Daniel E. Holz, and David N. Spergel. Limits on the number of spacetime dimensions from GW170817. 2018(7):048, July 2018.

[204] Rosalba Perna, Yi-Han Wang, Will M. Farr, Nathan Leigh, and Matteo Cantiello. Constraining the Black Hole Initial Mass Function with LIGO/VIRGO Observations. *arXiv e-prints*, page arXiv:1901.03345, Jan 2019.

[205] A. Petiteau, S. Babak, and A. Sesana. Constraining the Dark Energy Equation of State Using LISA Observations of Spinning Massive Black Hole Binaries. *ApJ*, 732:82, May 2011.

[206] E Sterl Phinney. The rate of neutron star binary mergers in the universe-minimal predictions for gravity wave detectors. *The Astrophysical Journal*, 380:L17–L21, 1991.

[207] M. H. Pinsonneault and K. Z. Stanek. Binaries Like to Be Twins: Implications for Doubly Degenerate Binaries, the Type Ia Supernova Rate, and Other Interacting Binaries. *ApJL*, 639:L67–L70, March 2006.

[208] M. Pürrer, M. Hannam, and F. Ohme. Can we measure individual black-hole spins from gravitational-wave observations? *Phys. Rev. D*, 93(8):084042, April 2016.

[209] G. Rakavy, G. Shaviv, and Z. Zinamon. Carbon and Oxygen Burning Stars and Pre-Supernova Models. *ApJ*, 150:131, October 1967.

[210] Carl L. Rodriguez, Pau Amaro-Seoane, Sourav Chatterjee, and Frederic A. Rasio. Post-Newtonian Dynamics in Dense Star Clusters: Highly Eccentric, Highly Spinning, and Repeated Binary Black Hole Mergers. *Phys. Rev. Lett.*, 120(15):151101, Apr 2018.

[211] Carl L. Rodriguez, Sourav Chatterjee, and Frederic A. Rasio. Binary black hole mergers from globular clusters: Masses, merger rates, and the impact of stellar evolution. *Phys. Rev. D*, 93(8):084029, Apr 2016.

[212] Carl L. Rodriguez, Carl-Johan Haster, Sourav Chatterjee, Vicky Kalogera, and Frederic A. Rasio. Dynamical Formation of the GW150914 Binary Black Hole. *ApJ*, 824(1):L8, Jun 2016.

[213] Carl L. Rodriguez and Abraham Loeb. Redshift Evolution of the Black Hole Merger Rate from Globular Clusters. *ApJ*, 866(1):L5, Oct 2018.

[214] Carl L. Rodriguez, Michael Zevin, Chris Pankow, Vasilliki Kalogera, and Frederic A. Rasio. Illuminating Black Hole Binary Formation Channels with Spins in Advanced LIGO. *ApJ*, 832(1):L2, Nov 2016.

[215] Javier Roulet and Matias Zaldarriaga. Constraints on binary black hole populations from LIGO-Virgo detections. *MNRAS*, 484(3):4216–4229, Apr 2019.

[216] E. E. Salpeter. The Luminosity Function and Stellar Evolution. *ApJ*, 121:161, January 1955.

[217] John Salvatier, Thomas V. Wiecki, and Christopher Fonnesbeck. Probabilistic programming in python using pymc3. *PeerJ Computer Science*, 2:e55, 2016.

[218] Johan Samsing. Eccentric black hole mergers forming in globular clusters. *Phys. Rev. D*, 97(10):103014, May 2018.

[219] B. F. Schutz. Determining the Hubble constant from gravitational wave observations. *Nature*, 323:310, September 1986.

[220] D. M. Scolnic, D. O. Jones, A. Rest, Y. C. Pan, R. Chornock, R. J. Foley, M. E. Huber, R. Kessler, G. Narayan, A. G. Riess, S. Rodney, E. Berger, P. J. Challis, M. Drout, D. Finkbeiner, R. Lunnan, R. P. Kirshner, N. E. Sanders, E. Schlafly, S. Smartt, C. W. Stubbs, J. Tonry, W. M. Wood-Vasey, M. Foley, J. Hand, E. Johnson, W. S. Burgett, K. C. Chambers, P. W. Draper, K. W. Hodapp, N. Kaiser, R. P. Kudritzki, E. A. Magnier, N. Metcalfe, F. Bresolin, E. Gall, R. Kotak, M. McCrum, and K. W. Smith. The Complete Light-curve Sample of Spectroscopically Confirmed Type Ia Supernovae from Pan-STARRS1 and Cosmological Constraints from The Combined Pantheon Sample. *ArXiv e-prints*, October 2017.

[221] Claude Shannon. A mathematical theory of communication. *Bell Labs Technical Journal*, 27:379–423, 07 1948.

[222] S. Sigurdsson and L. Hernquist. Primordial black holes in globular clusters. *Nature*, 364:423–425, July 1993.

[223] L. P. Singer, H.-Y. Chen, D. E. Holz, W. M. Farr, L. R. Price, V. Raymond, S. B. Cenko, N. Gehrels, J. Cannizzo, M. M. Kasliwal, S. Nissanke, M. Coughlin, B. Farr, A. L. Urban, S. Vitale, J. Veitch, P. Graff, C. P. L. Berry, S. Mohapatra, and I. Mandel. Going the Distance: Mapping Host Galaxies of LIGO and Virgo Sources in Three Dimensions Using Local Cosmography and Targeted Follow-up. *ApJL*, 829:L15, September 2016.

[224] L. P. Singer, H.-Y. Chen, D. E. Holz, W. M. Farr, L. R. Price, V. Raymond, S. B. Cenko, N. Gehrels, J. Cannizzo, M. M. Kasliwal, S. Nissanke, M. Coughlin, B. Farr, A. L. Urban, S. Vitale, J. Veitch, P. Graff, C. P. L. Berry, S. Mohapatra, and I. Mandel. Supplement: Going the Distance: Mapping Host Galaxies of LIGO and Virgo Sources in Three Dimensions Using Local Cosmography and Targeted Follow-up (2016, ApJL, 829, L15). *apjs*, 226:10, September 2016.

[225] Leo P. Singer, Larry R. Price, Ben Farr, Alex L. Urban, Chris Pankow, Salvatore Vitale, John Veitch, Will M. Farr, Chad Hanna, Kipp Cannon, Tom Downes, Philip Graff, Carl-Johan Haster, Ilya Mandel, Trevor Sidery, and Alberto Vecchio. The First Two Years of Electromagnetic Follow-up with Advanced LIGO and Virgo. *ApJ*, 795:105, November 2014.

[226] J. Slutsky, L. Blackburn, D. A. Brown, L. Cadonati, J. Cain, M. Cavaglià, S. Chatterji, N. Christensen, M. Coughlin, S. Desai, G. González, T. Isogai, E. Katsavounidis, B. Rankins, T. Reed, K. Riles, P. Shawhan, J. R. Smith, N. Zotov, and J. Zweizig. Methods for reducing false alarms in searches for compact binary coalescences in LIGO data. *Classical and Quantum Gravity*, 27(16):165023, August 2010.

[227] N. Smirnov. Table for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics*, 19(2):279281, 1948.

[228] Graham P. Smith, Christopher Berry, Matteo Bianconi, Will M. Farr, Mathilde Jauzac, Richard Massey, Johan Richard, Andrew Robertson, Keren Sharon, Alberto Vecchio, and John Veitch. Strong-lensing of Gravitational Waves by Galaxy Clusters. In G. González and R. Hynes, editors, *IAU Symposium*, volume 338 of *IAU Symposium*, pages 98–102, October 2018.

[229] M. Soares-Santos, D. E. Holz, J. Annis, R. Chornock, K. Herner, Berger, and et al. The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. I. Discovery of the Optical Counterpart Using the Dark Energy Camera. *ApJL*, 848:L16, October 2017.

[230] M. Spera, N. Giacobbo, and M. Mapelli. Shedding light on the black hole mass spectrum. *MmSAI*, 87:575, 2016.

[231] M. Spera and M. Mapelli. Very massive stars, pair-instability supernovae and intermediate-mass black holes with the sevn code. *MNRAS*, 470:4739–4749, October 2017.

[232] Mario Spera, Michela Mapelli, and Alessandro Bressan. The mass spectrum of compact remnants from the PARSEC stellar evolution tracks. *MNRAS*, 451(4):4086–4103, Aug 2015.

[233] Mario Spera, Michela Mapelli, Nicola Giacobbo, Alessandro A. Trani, Alessandro Bressan, and Guglielmo Costa. Merging black hole binaries with the SEVN code. *MNRAS*, 485(1):889–907, May 2019.

[234] Mario Spera, Michela Mapelli, Nicola Giacobbo, Alessandro A. Trani, Alessandro Bressan, and Guglielmo Costa. Merging black hole binaries with the SEVN code. *MNRAS*, 485(1):889–907, May 2019.

[235] M. A. Stephens. EDF statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association*, 69(347):730–737, September 1974.

[236] S. Stevenson, C. P. L. Berry, and I. Mandel. Hierarchical analysis of gravitational-wave measurements of binary black hole spin-orbit misalignments. *ArXiv e-prints*, March 2017.

[237] Simon Stevenson, Frank Ohme, and Stephen Fairhurst. Distinguishing Compact Binary Population Synthesis Models Using Gravitational Wave Observations of Coalescing Binary Black Holes. *ApJ*, 810(1):58, Sep 2015.

[238] Simon Stevenson, Alejandro Vigna-Gómez, Ilya Mandel, Jim W. Barrett, Coenraad J. Neijssel, David Perkins, and Selma E. de Mink. Formation of the first three gravitational-wave observations through isolated binary evolution. *Nature Communications*, 8:14906 EP –, 04 2017.

[239] Louis E. Strigari, John F. Beacom, Terry P. Walker, and Pengjie Zhang. The concordance cosmic star formation rate: implications from and for the supernova neutrino and gamma ray backgrounds. 2005(4):017, April 2005.

[240] Colm Talbot and Eric Thrane. Measuring the Binary Black Hole Mass Spectrum with an Astrophysically Motivated Parameterization. *ApJ*, 856(2):173, Apr 2018.

[241] Stephen R. Taylor and Davide Gerosa. Mining gravitational-wave catalogs to understand binary stellar evolution: A new hierarchical Bayesian framework. *Phys. Rev. D*, 98(8):083017, Oct 2018.

[242] W. Tichy and P. Marronetti. Final mass and spin of black-hole mergers. *Phys. Rev. D*, 78(8):081501, October 2008.

[243] S. A. Usman, A. H. Nitz, I. W. Harry, C. M. Biwer, D. A. Brown, M. Cabero, C. D. Capano, T. Dal Canton, T. Dent, S. Fairhurst, M. S. Kehl, D. Keppel, B. Krishnan, A. Lenon, A. Lundgren, A. B. Nielsen, L. P. Pekowsky, H. P. Pfeiffer, P. R. Saulson, M. West, and J. L. Willis. The PyCBC search for gravitational waves from compact binary coalescence. *Classical and Quantum Gravity*, 33(21):215004, November 2016.

[244] Elisabeth Vangioni, Keith A. Olive, Tanner Prestegard, Joseph Silk, Patrick Petitjean, and Vuk Mandic. The impact of star formation and gamma-ray burst rates at high redshift on cosmic chemical evolution and reionization. *MNRAS*, 447(3):2575–2587, March 2015.

[245] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, W. Del Pozzo, F. Feroz, J. Gair, C.-J. Haster, V. Kalogera, T. Littenberg, I. Mandel, R. O'Shaughnessy, M. Pitkin, C. Rodriguez, C. Röver, T. Sidery, R. Smith, M. Van Der Sluys, A. Vecchio, W. Vousden, and L. Wade. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Phys. Rev. D*, 91(4):042003, February 2015.

[246] John Veitch, Michael Pürrer, and Ilya Mandel. Measuring intermediate-mass black-hole binaries with advanced gravitational wave detectors. *Phys. Rev. Lett.*, 115:141101, Sep 2015.

[247] Tejaswi Venumadhav, Barak Zackay, Javier Roulet, Liang Dai, and Matias Zaldarriaga. New Binary Black Hole Mergers in the Second Observing Run of Advanced LIGO and Advanced Virgo. *arXiv e-prints*, page arXiv:1904.07214, Apr 2019.

[248] Tejaswi Venumadhav, Barak Zackay, Javier Roulet, Liang Dai, and Matias Zaldarriaga. New search pipeline for compact binary mergers: Results for binary black holes in the first observing run of Advanced LIGO. *Phys. Rev. D*, 100(2):023011, Jul 2019.

[249] S. Vitale, R. Lynch, J. Veitch, V. Raymond, and R. Sturani. Measuring the Spin of Black Holes in Binary Systems Using Gravitational Waves. *Physical Review Letters*, 112(25):251101, June 2014.

[250] Salvatore Vitale and Will M. Farr. Measuring the star formation rate with gravitational waves from binary black holes. *arXiv e-prints*, page arXiv:1808.00901, Aug 2018.

[251] Salvatore Vitale, Ryan Lynch, Vivien Raymond, Riccardo Sturani, John Veitch, and Philip Graff. Parameter estimation for heavy binary-black holes with networks of second-generation gravitational-wave detectors. *Phys. Rev. D*, 95:064053, Mar 2017.

[252] Salvatore Vitale, Ryan Lynch, Riccardo Sturani, and Philip Graff. Use of gravitational waves to probe the formation channels of compact binaries. *Classical and Quantum Gravity*, 34(3):03LT01, Feb 2017.

[253] S. E. Woosley. The Progenitor of GW150914. *ApJ*, 824(1):L10, Jun 2016.

[254] S. E. Woosley. Pulsational Pair-instability Supernovae. *ApJ*, 836:244, February 2017.

[255] Daniel Wysocki, Jacob Lange, and Richard O'Shaughnessy. Reconstructing phenomenological distributions of compact binaries via gravitational wave observations. *Phys. Rev. D*, 100(4):043012, Aug 2019.

[256] Michael Zevin, Chris Pankow, Carl L. Rodriguez, Laura Sampson, Eve Chase, Vassiliki Kalogera, and Frederic A. Rasio. Constraining Formation Models of Binary Black Holes with Gravitational-wave Observations. *ApJ*, 846(1):82, Sep 2017.

[257] Michael Zevin, Johan Samsing, Carl Rodriguez, Carl-Johan Haster, and Enrico Ramirez-Ruiz. Eccentric Black Hole Mergers in Dense Star Clusters: The Role of Binary-Binary Encounters. *ApJ*, 871(1):91, Jan 2019.