Where Are LIGO’s Big Black Holes?

Maya Fishbach1 and Daniel E. Holz2

1 Department of Astronomy and Astrophysics, University of Chicago, Chicago, IL 60637, USA
2 Enrico Fermi Institute, Department of Physics, Department of Astronomy and Astrophysics, and Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA

Received 2017 October 11; revised 2017 November 15; accepted 2017 November 17; published 2017 December 11

Abstract

In LIGO’s O1 and O2 observational runs, the detectors were sensitive to stellar-mass binary black hole (BBH) coalescences with component masses up to 100 $M_\odot$, with binaries with primary masses above 40 $M_\odot$, representing $\geq 90\%$ of the total accessible sensitive volume. Nonetheless, of the 5.9 detections (GW150914, LVT151012, GW151226, GW170104, GW170608, and 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 BBHs. 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 $M_{\text{max}} \sim 40 M_\odot$. This inferred distribution is entirely consistent with the two recently reported detections, GW170608 and GW170814. We show that with $\sim 10$ additional LIGO-Virgo BBH detections, fitting the BBH mass distribution will provide strong evidence for an upper mass gap if one exists.

Key words: binaries: general – gravitational waves – methods: data analysis – stars: black holes – stars: massive

1. Introduction

One of the most fundamental quantities in gravitational-wave (GW) 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 detections of binary black holes (BBHs), the sample of $\sim 20$ BHs in X-ray binary systems was used to infer the BH mass distribution (Özel et al. 2010; Farr et al. 2011), providing strong evidence for the existence of a mass gap between the heaviest neutron star (NS: $\sim 2$–3 $M_\odot$) and the lightest BH ($\sim 4$–5 $M_\odot$; but see also Keidberg et al. 2012). The presence of a mass gap between NSs and BHs has critical implications for supernova explosion theory (Belczynski et al. 2012), and there are several proposed methods to probe this mass gap with GW observations of compact binaries (Littenberg et al. 2015; Kovetz et al. 2017; Mandel et al. 2017). In addition to the low-mass gap, supernova theory suggests that pulsational pair-instability supernovae (PPISN; Heger & Woosley 2002) and pair-instability supernovae (PISN; Fowler & Hoyle 1964; Rakavy et al. 1967; Bond et al. 1984) lead to a second mass gap between $\sim 50$ and 135 $M_\odot$ for BHs formed from stellar core collapse (Belczynski et al. 2014, 2016a; Marchant et al. 2016; Spera & Mapelli 2017; Woosley 2017). Although studies of the lower-mass gap have to wait for more binary detections, because LIGO’s sensitivity is almost 500 times greater for 50–50 $M_\odot$ mergers than 3–3 $M_\odot$ mergers, the existing data already begin 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 parameterization on the primary BBH mass, $m_1$ (Abbott et al. 2016a, 2017b), inspired by the stellar initial mass function (IMF) (Salpeter 1955; Kroupa 2001). Specifically, Abbott et al. (2016a) use the following one-parameter power law to model the distribution of primary component BH masses:

$$p(m_1|\alpha) \propto m_1^{-\alpha},$$ (1)

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

$$p(m_2|m_1) \propto \frac{1}{\min(m_1, M_{\text{tot,max}} - m_1) - M_{\text{min}}}.$$(2)

and therefore the joint mass distribution is

$$p(m_1, m_2|\alpha) \propto \frac{m_1^{-\alpha}}{\min(m_1, M_{\text{tot,max}} - m_1) - M_{\text{min}}}.$$ (3)

The only free parameter in this assumed mass distribution is the power-law slope, $\alpha$. The minimum BH mass, $M_{\text{min}}$, is fixed at $M_{\text{min}} = 5 M_\odot$, and the maximum mass, $M_{\text{max}}$, is fixed at $M_{\text{max}} = 100 M_\odot$. Meanwhile, the total BBH mass, $M_{\text{tot}} = m_1 + m_2$, is also restricted: $M_{\text{tot}} \leq M_{\text{tot,max}} = 100 M_\odot$. (Enforcing $M_{\text{tot,max}} = 100 M_\odot$ causes a break in the power law at $50 M_\odot$.) The choice for $M_{\text{min}}$ is motivated by the empirical lower-mass gap, while $M_{\text{max}}$ and $M_{\text{tot,max}}$ are set by the stellar binary matched-filter search, which defines stellar-mass BBHs as those with source-frame total masses $m_1 + m_2 \leq 100 M_\odot$ (Cannon et al. 2012; Abbott et al. 2016b; Usman et al. 2016). However, LIGO is in principle sensitive to heavier BHs (Abbott et al. 2017a). 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 (Dal Canton & Harry 2017; Messick et al. 2017; Nitz et al. 2017), and IMBHs of even higher mass can be detected by the
unmodeled transient search (Klimenko et al. 2008; Abbott et al. 2017a).

We observe that a key assumption of the distribution in Equation (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 BHs 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 $10 M_\odot < M_{\text{tot}} < 100 M_\odot$, it is still a good approximation to take $VT \propto m_1^{5}$, with $k \sim 2.2$. This means that if the BBH mass distribution follows a power law with slope $\alpha$ as in Equation (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 BBH masses. Thus, the absence of heavy BBHs in the data quickly indicates either that the mass distribution declines steeply toward 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 LVTV151012, which has an 87% probability of being astrophysical; Abbott et al. 2016a, 2017b). 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 that 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. 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; Abbott et al. 2017d, 2017c). We show that with $O(10)$ additional detections, the presence and location of the bottom edge of the mass gap will be highly constrained.

2. 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 (Abbott et al. 2016d). Under these assumptions, $VT$ is given by

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

(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 Ade et al. (2016) throughout the calculation. To calculate the detection probability, $f(z, m_1, m_2)$, we use the semi-analytic approximation from Abbott et al. (2016d). Taking the PSD function corresponding to the early aLIGO high-sensitivity scenario in Abbott et al. (2016c; a good approximation to the PSD during the first and second aLIGO observing runs), we calculate the optimal matched-filter signal-to-noise ratio ($S/N$) of a BBH with component masses $m_1$ and $m_2$ and zero spins located at redshift $z$. The optimal $S/N$, $\rho_{\text{opt}}$, corresponds to a face-on source that is directly overhead to a single detector. We then generate random angular factors, $0 < \psi < 1$, from a single-detector antenna power pattern, assuming that binaries are distributed uniformly on the sky with isotropic inclination vectors (Finn & Chernoff 1993; Dominik et al. 2015; Belczynski et al. 2016c). The angular factor, $\psi$, characterizes the response of a detector to a source at a given sky location and orientation (so that $\psi = 1$ for an overhead, face-on source). From the distribution of angular factors, $\psi$, we assign a distribution of $S/N_s$, $\rho = \rho_{\text{opt}}$, for each source with parameters $(m_1, m_2, z)$. Out of this distribution of $S/N_s$, the fraction that exceeds 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 $S/N$ 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 $M_{\text{tot}} = 100 M_\odot$ and possibly higher (Abbott et al. 2016d, 2017e).

The expected sensitive redshifted spacetime volume as a function of total mass is shown in Figure 1, 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 $m_1 = M_{\text{tot}}/(1 + q)$ for a fixed mass ratio $q$, we can see from Figure 1 that $VT$ scales approximately as $m_1^{2}$ with $k \sim 2.2$ for $M_{\text{tot}} \lesssim 100 M_\odot$.

To calculate the sensitivity to a population of BBHs, the relevant quantity is the population-averaged spacetime volume,
If we know the distribution of masses across the population of BBHs, $p_{\text{pop}}(m_1, m_2)$, assuming negligible spins and a constant comoving merger rate, we can calculate the population-averaged sensitive spacetime volume (Equation (15) in Abbott et al. 2016d):

$$\langle VT \rangle = \int \int VT(m_1, m_2)p_{\text{pop}}(m_1, m_2)dm_1dm_2,$$

where the first integral is over $M_{\text{min}} < m_1 < M_{\text{max}}$ and the second integral is over $M_{\text{min}} < m_2 < \min(m_1, M_{\text{max}})$. \langle VT \rangle relates the specific merger rate, $R$, to the expected number, $\Lambda$, of BBH signals in a given detection period (Abbott et al. 2016d):

$$\Lambda = R \langle VT \rangle.$$  

(6)

The number of BBH detections, $n$, follows a Poisson process with mean $\Lambda$. To explore the existence of a high-mass gap in Section 4.1, we compare the expected number of low-mass BBH signals, $\Lambda_{\text{low}} = R \langle VT \rangle_{\text{low}}$, to the expected number of high-mass BBH signals, $\Lambda_{\text{high}} = R \langle VT \rangle_{\text{high}}$, for different power-law populations, where low (high) mass is defined by the primary component mass $m_1 \leq M_{\text{cutoff}}$ ($m_1 > M_{\text{cutoff}}$). We define

$$\frac{1}{r} = \frac{\Lambda_{\text{low}}}{\Lambda_{\text{high}}} = \frac{\langle VT \rangle_{\text{low}}}{\langle VT \rangle_{\text{high}}},$$

(7)

where

$$\langle VT \rangle_{\text{low}} = \int \int_{M_{\text{min}}}^{M_{\text{cutoff}}} VT(m_1, m_2)p_{\text{pop}}(m_1, m_2)dm_1dm_2$$

$$\langle VT \rangle_{\text{high}} = \int \int_{M_{\text{cutoff}}}^{M_{\text{max}}} VT(m_1, m_2)p_{\text{pop}}(m_1, m_2)dm_1dm_2.$$

(8)

The integration limits on the $m_2$ integral in Equation (8) are identical to those in Equation (5) so that the total $\langle VT \rangle = \langle VT \rangle_{\text{low}} + \langle VT \rangle_{\text{high}}$. We can then compute the probability of detecting $n_{\text{high}}$ BBHs with primary component mass $m_1 > M_{\text{cutoff}}$, given that we have detected $n_{\text{low}}$ BBHs with primary component mass $m_1 < M_{\text{cutoff}}$. (We ignore mass measurement uncertainties that may prevent us from definitively assigning a BBH to either the low- or high-mass class.) This probability is given by

$$p(n_{\text{high}}|n_{\text{low}}) = \int_0^\infty \int_0^\infty p(n_{\text{high}}, n_{\text{low}}|\Lambda_{\text{high}}, \Lambda_{\text{low}})$$

$$\times (\Lambda_{\text{high}}|n_{\text{high}})d\Lambda_{\text{high}}d\Lambda_{\text{low}}$$

$$= \int \int p(n_{\text{high}}|\Lambda_{\text{high}})p(\Lambda_{\text{high}}|\Lambda_{\text{low}})p$$

$$\times (\Lambda_{\text{low}}|n_{\text{low}})d\Lambda_{\text{high}}d\Lambda_{\text{low}}$$

$$= \int \int p(n_{\text{high}}|\Lambda_{\text{high}})\delta(\Lambda_{\text{high}} - r\Lambda_{\text{low}})p$$

$$\times (\Lambda_{\text{low}}|n_{\text{low}})d\Lambda_{\text{high}}d\Lambda_{\text{low}}$$

$$= \int p(n_{\text{high}}|r\Lambda_{\text{low}})p(\Lambda_{\text{low}}|n_{\text{low}})d\Lambda_{\text{low}}$$

$$\propto \int p(n_{\text{high}}|r\Lambda_{\text{low}})p(n_{\text{low}}|\Lambda_{\text{low}})p_0(\Lambda_{\text{low}})d\Lambda_{\text{low}},$$

(9)

where in the third line we used the definition of $r$ given by Equation (8) and in the last line we used Bayes's theorem. In Equation (9), terms like $p(n|\Lambda)$ denote the Poisson probability of $n$ with mean $\Lambda$. We take the prior $p_0(\Lambda_{\text{low}})$ to be the Jeffrey’s prior:

$$p_0(\Lambda_{\text{low}}) \propto \frac{1}{\sqrt{\Lambda_{\text{low}}}}.$$  

(10)

We will return to Equation (9) in Section 4.1.

### 3. 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, $M_{\text{max}}$. We therefore follow Abbott et al. (2016a) in fitting a power-law mass distribution to GW BBH mass measurements, but we add the maximum BH mass, $M_{\text{max}}$, as a free parameter. We leave the minimum BH mass, $M_{\text{min}}$, fixed at $M_{\text{min}} = 5M_\odot$. Thus, we consider a two-parameter mass distribution:

$$p(m_1, m_2|\alpha, M_{\text{max}}) \propto \frac{m_1^{-\alpha}H(M_{\text{max}} - m_1)}{\min(m_1, M_{\text{max}} - m_1) - M_{\text{min}}},$$

(11)

where $H$ is the Heaviside step function that enforces a cutoff in the distribution at $m_1 = M_{\text{max}}$. For consistency with the LIGO definition of a stellar-mass BH, we restrict $M_{\text{max}} \leq 100M_\odot$ throughout. Furthermore, as in the LIGO Collaboration’s analysis, we enforce $m_1 + m_2 \leq M_{\text{max}}$, which provides an additional constraint for $M_{\text{max}} > \frac{1}{2}M_{\text{tot,max}}$. The LIGO Collaboration fixes $M_{\text{tot,max}} = 100M_\odot$ and $M_{\text{min}} = 100M_\odot - M_{\text{min}}$, 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 $M_{\text{tot}} > 100M_\odot$ 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 $M_{\text{tot}} > 100M_\odot$, so we cannot constrain their existence. (In the absence of detections with $M_{\text{tot,max}} > 100M_\odot$, setting $M_{\text{tot,max}} \leq 100M_\odot$ in the population model, Equation (11), is equivalent to assuming that the sensitivity vanishes for binaries with $M_{\text{tot}} > 100M_\odot$.) Assumption (a) may not be well motivated, as population-synthesis models that predict stellar BBHs with component masses $M_{\text{max}} \sim 50$--$100M_\odot$ tend to allow $M_{\text{tot,max}} \sim 2M_{\text{max}}$ (Belczynski et al. 2016b; Eldridge & Stanway 2016). Assumption (b) can also be questioned, as LIGO is sensitive to BBHs with detector-frame total masses up to at least $600M_\odot$ in the IMBH matched-filter search (Abbott et al. 2017e). However, the sensitivity may be lower than expected for very high mass BBHs due to non-stationary instrumental noise (Slutsky et al. 2010) 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 (Capano et al. 2014; Calderón Bustillo et al. 2017a, 2017b). If we had an accurate model of $VT(m_1, m_2)$ across the mass range $5 < m_1, m_2 < 100M_\odot$ (by performing a large-scale injection campaign), we could set $M_{\text{tot,max}} = 2M_{\text{max}} \leq 200M_\odot$ in Equation (11). However, because our calculation of $VT$ may be overestimating the sensitivity to binaries with $M_{\text{tot,max}} > 100M_\odot$, when fitting Equation (11) in the following sections, we repeat the analysis once under the assumption that...
\[ M_{\text{tot}, \text{max}} = \min(2M_{\text{max}}, 100M_\odot) \] and once assuming that \[ M_{\text{tot}, \text{max}} = 2M_{\text{max}} \leq 200M_\odot. \]

To extract the parameters of our assumed mass distribution (Equation (11)) from data, we use the same hierarchical Bayesian methods as Appendix D of Abbott et al. (2016a), further explained in Mandel et al. (2016). While GW data are 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 2 and Abbott et al. (2016d).

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, \( d \), from a BBH detection (Veitch et al. 2015). 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, \( p(m_1, m_2|d) \). 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, and 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 Abbott et al. (2016a) or Table 1 of Abbott et al. (2017b). 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 S/N from the antenna power pattern, using the PSD corresponding to the early aLIGO high-sensitivity scenario (as described in Section 2). Out of this population, the set of detections are those simulated BBHs with a single-detector S/N satisfying \( \rho > 8 \). Given the true component masses and the S/N of each mock BBH detection, we produce realistic mass measurements by generating 5000–10,000 posterior samples for the component masses following the prescription in Equation (1) of Mandel et al. (2017). Given true values for the chirp mass, \( M^\star \), symmetric mass ratio, \( \eta^\star \), and S/N, \( \rho^\star \), we draw chirp mass posterior samples from a Gaussian distribution centered at \( M^\star \) with standard deviation \( \sigma_M \) and symmetric mass ratio posterior samples from a Gaussian distribution centered at \( \eta^\star \) with standard deviation \( \sigma_\eta \). We only keep posterior samples with \( 0.01 \leq \eta \leq 0.25 \). The variables \( \hat{M} \) and \( \hat{\eta} \) are drawn from Gaussian distributions:

\[
\begin{align*}
\hat{M} &\sim N(M^\star, \sigma_M), \\
\hat{\eta} &\sim N(\eta^\star, \sigma_\eta),
\end{align*}
\]

where \( \sigma_M, \sigma_\eta \) scale inversely with the S/N and are given in Mandel et al. (2017).

Once we have samples from the posterior PDF, \( p(m_1, m_2|d) \), for each event (both real and simulated) and we have calculated the detection efficiency, \( \mathcal{P}_{\text{det}}(m_1, m_2) \propto VT(m_1, m_2) \), we follow Appendix D in Abbott et al. (2016a) to fit Equation (11). The likelihood for a single BBH detection given the parameters of the mass distribution, \( \alpha \) and \( M_{\text{max}} \), is given by

\[
p(d|\alpha, M_{\text{max}}) \propto \int p(d|m_1, m_2)p(m_1, m_2|\alpha, M_{\text{max}})dm_1dm_2
\]

\[
\propto \left( \frac{\beta(\alpha, M_{\text{max}})}{\beta(\alpha, M_{\text{max}})} \right),
\]

where \( \langle \ldots \rangle \) denotes an average over the \( (m_1, m_2) \) posterior samples. This is valid because for each event, \( p(d|m_1, m_2) \propto p(m_1, m_2|d) \), as the prior on \( m_1, m_2 \) is taken to be flat. Therefore, we can calculate the integral in the first line of Equation (13) by taking the average of \( p(m_1, m_2|\alpha, M_{\text{max}}) \) over the mass posterior samples. Meanwhile, \( \beta(\alpha, M_{\text{max}}) \) is defined as

\[
\beta(\alpha, M_{\text{max}}) \equiv \int p(m_1, m_2|\alpha, M_{\text{max}})VT(m_1, m_2)dm_1dm_2.
\]

The likelihood for the data across all events \( d = \{d\} \) is the product of the individual event likelihoods given by Equation (13).

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

\[
p(d|M_{\text{max}}) = \int p(d|\alpha, M_{\text{max}})p_0(\alpha|M_{\text{max}})d\alpha.
\]

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

\[
K(M, 100M_\odot) = \prod_{i=1}^{N} \frac{p(d|M_{\text{max}} = M)}{p(d|M_{\text{max}} = 100M_\odot)}.
\]

We calculate the cumulative Bayes factor \( K(M, 100M_\odot) \) in Section 4.2.

4. Results

4.1. Non-detection of Heavy BBHs

We first give an example of how the detection of only four BBHs with primary masses \( m \leq 40–50M_\odot \) is inconsistent with certain (possibly correct) power-law mass distributions.
unless a mass gap is imposed. For a given mass distribution, we can use Equation (9) to calculate the probability of not detecting any BBHs above a certain mass, \( n_{\text{low}} \), given that we have detected \( n_{\text{high}} \) BBHs below the cutoff mass. To do this, we must first compute the ratio \( \frac{\langle VT \rangle_{\text{low}}}{\langle VT \rangle_{\text{high}}} \) as defined in Equation (7) and then apply Equation (9). The results of this calculation for \( \frac{\langle VT \rangle_{\text{low}}}{\langle VT \rangle_{\text{high}}} = 0 \) are displayed in Figure 2. We show the results for two choices of cutoff mass: \( M_{\text{cut}} = 41 M_\odot \) (green curve) is motivated by the 95% credible upper limit on the primary mass of GW150914, the heaviest BBH detected, and \( M_{\text{cut}} = 50 M_\odot \) (blue and orange curves) is motivated by PPSN and PISN supernova models, which predict a mass gap starting at \( 40 \)–\( 50 M_\odot \) (depending also on details of binary evolution; Belczynski et al. 2016a; Woosley 2017). We also vary the maximum total mass, \( M_{\text{tot,max}} \), of the “high-mass” population between \( 100 M_\odot \) (blue and green curves), which is currently 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 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 \( M_{\text{tot,max}} = 100 M_\odot \), or at the \( >99.9\% \) level if we assume that the BBH population and the detectors’ sensitivity extends up to \( M_{\text{tot,max}} = 200 M_\odot \). 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 \( M_{\text{tot,max}} = 200 M_\odot \). 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.

### 4.2. 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 \( M_{\text{max}} = 41 M_\odot \) (Equation (11), but we follow the canonical analysis used by the LVC (see Equation (3)) and fix \( M_{\text{max}} = 95 M_\odot \) and \( M_{\text{tot,max}} = 100 M_\odot \) when inferring the power-law slope. While the bias on the inferred slope \( \alpha \) 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 Figure 3). (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 \( M_{\text{max}} = 100 M_\odot \) and \( M_{\text{tot,max}} = 200 M_\odot \) rather than \( M_{\text{tot,max}} = 100 M_\odot \), the presence of a mass gap will bias the power-law inference even more significantly, as the true population has \( M_{\text{tot,max}} = 82 M_\odot \). 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 parameterized fit to the BBH mass distribution.

With the first four LIGO BBH detections, varying \( M_{\text{max}} \) does not drastically bias the inference on \( \alpha \) when fitting Equation (3) (see the solid lines in Figure 3). However, we can distinguish the model favored by the data by calculating the cumulative Bayes factor. Following Equations (15)–(17), we calculate this factor between two single-parameter power-law models with different fixed values of \( M_{\text{max}} \). 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 probability of \( M_{\text{max}} \) in Equation (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 \( M_{\text{tot,max}} = 100 M_\odot \) (so that \( M_{\text{max}} = 100 M_\odot \) is really \( M_{\text{max}} = 95 M_\odot \). If we instead assume that the \( M_{\text{max}} = 100 M_\odot \) population is fully detectable up to total binary masses of \( M_{\text{tot,max}} = 200 M_\odot \), 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 \( M_{\text{max}} \sim 40 M_\odot \), over a cutoff at \( M_{\text{max}} \sim 100 M_\odot \) within the assumed power-law model (Kass & Raftery 1995). The Bayes factor also depends on the choice of prior on \( \alpha \). 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 (\( -2 < \alpha < 0 \)). It is clear from Figure 2 that a prior that favors large positive values of \( \alpha \) (steeply declining power-law slopes) will lower the evidence in favor of a mass cutoff \( M_{\text{max}} > 100 M_\odot \), while placing greater prior support on low values of \( \alpha \) (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 \( K(41 M_\odot, 100 M_\odot) = 5 \).
$M_{\text{tot, max}} \leq 100 M_{\odot}$ or $K(41 M_{\odot}, 100 M_{\odot}) = 21$ if we allow $M_{\text{tot, max}} \leq 200 M_{\odot}$.

We anticipate that a set of 10 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 $2M_{\text{max}}$ in either case, so $M_{\text{tot, max}} \leq 200 M_{\odot}$. We take a flat prior on $\alpha$ in the range $[-2, 7]$. With 191 events from a simulated $\alpha = 2.35$, $M_{\text{max}} = 41 M_{\odot}$ 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 10 BBH detections from this population, we get $K(41 M_{\odot}, 100 M_{\odot}) > 150$ in more than 99% of cases. If we detect 10 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 $M_{\text{max}}$ favored by a given set of detections in Section 4.3.

**Figure 3.** Inferred likelihood for the power-law slope of the mass distribution, $\alpha$, calculated for 120 mock observations from a $M_{\text{max}} = 41 M_{\odot}$, $\alpha = 2.35$ 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 $M_{\text{max}} = 95 M_{\odot}$, while the pink and green curves correspond to a fixed maximum mass at $M_{\text{max}} = 41 M_{\odot}$. 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.

4.3. Joint Power Law–Maximum Mass Fit

In this section, we fit the two-parameter mass distribution of Equation (11). We calculate the likelihood $p(d|\alpha, M_{\text{max}})$ as the product of individual event likelihoods in Equation (13). We take flat priors on $\alpha$ and $M_{\text{max}}$ with $-2 \leq \alpha \leq 7$ as before and $M_{\text{max}} \leq 100 M_{\odot}$. The minimum allowed value of $M_{\text{max}}$ 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 $M_{\text{max}} \leq 100 M_{\odot}$ 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 (Belczynski et al. 2016a; Eldridge & Stanway 2016; Spera et al. 2016). We calculate the likelihood function on a $500 \times 100$ grid of $(\alpha, M_{\text{max}})$ values in the allowed prior range and verify that increasing the resolution of the $(\alpha, M_{\text{max}})$ grid does not change our results. In fact, the resolution in the $M_{\text{max}}$ 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 $M_{\text{max}}$ 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 Figure 4. We compute the joint likelihood twice: once fixing $M_{\text{tot, max}} = 100 M_{\odot}$, so that the population of stellar-mass BBHs is restricted to total masses $M_{\text{tot}} \leq 100 M_{\odot}$ regardless of $M_{\text{max}}$ (top right panel) and once fixing $M_{\text{tot, max}} = 200 M_{\odot}$, so that the maximum total mass of the population is allowed to extend to $M_{\text{tot, max}} = 2M_{\text{max}}$. We calculate the marginal posterior PDFs of $\alpha$ and $M_{\text{max}}$ (top left and bottom right panels) under the assumption of a uniform prior on $\alpha$ in the range $[-2, 7]$ and a uniform prior on $M_{\text{max}}$ 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 $M_{\text{max}}$ when fitting the power-law mass distribution increases the support for shallow power-law slopes that would otherwise be ruled out under the assumption that the mass distribution extends continuously to $\sim 100 M_{\odot}$. Allowing for freedom in $M_{\text{max}}$ shifts the preferred values of $\alpha$ to shallower slopes, even allowing for negative $\alpha$, as compared to the canonical analysis that fixes $M_{\text{max}} = 95 M_{\odot}$ (orange dashed–dashed curve in Figure 4). Furthermore, the first four BBH detections already start to constrain $M_{\text{max}}$. The marginal posterior PDF $p(M_{\text{max}})$ peaks strongly at $M_{\text{max}} \sim 40$, and the 95% upper limits on the inferred $p(M_{\text{max}})$ are $76.6 M_{\odot}$ if assuming $M_{\text{tot, max}} = 200 M_{\odot}$ (or $90.7 M_{\odot}$ if we conservatively assume $M_{\text{tot, max}} = 100 M_{\odot}$). Taking $M_{\text{tot, max}} = 200 M_{\odot}$ rather than $100 M_{\odot}$ allows the detectable BBH population to extend to $2M_{\text{max}}$, thereby increasing the expected sensitivity to BBHs with primary component masses $M_{\text{max}} > 50 M_{\odot}$. Thus, the non-detection of heavy BBHs yields tighter constraints on the inferred $M_{\text{max}}$ when we assume $M_{\text{tot, max}} = 200 M_{\odot}$, but the peak of the $M_{\text{max}}$ 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 $M_{\text{max}} = 50 M_{\odot}$, while the other $\alpha = 2.35$ population has a mass gap starting at $M_{\text{max}} = 40 M_{\odot}$.

The results of this calculation, assuming sensitivity up to $M_{\text{tot, max}} = 200 M_{\odot}$, are shown in Figure 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 $M_{\text{max}} = 40 M_{\odot}$ (right column in Figure 5) rather than 50 $M_{\odot}$, we get tighter constraints on $M_{\text{max}}$. 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 Figure 2). 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

**Figure 4.** Marginal posterior PDFs of $\alpha$ and $M_{\text{max}}$ for the first four detected BBHs. The solid, dashed, and dot-dashed curves correspond to the canonical analysis with uniform priors on $\alpha$ and $M_{\text{max}}$, the purple and green curves correspond to a uniform prior on $\alpha$ and a fixed $M_{\text{max}} = 95 M_{\odot}$, while the pink and orange curves correspond to a fixed $M_{\text{max}} = 41 M_{\odot}$. The orange dashed–dashed curve in the top left panel corresponds to the canonical analysis with $M_{\text{tot, max}} = 200 M_{\odot}$.
but steeper power-law slope $\alpha = 2.35$ (left column). As the first four LIGO detections currently favor $M_{\text{max}} < 50 M_\odot$ and moderately shallow power-law slopes $\alpha < 3$ (Figure 4), we expect that fewer than 40 detections will strongly constrain $M_{\text{max}}$.

$M_{\text{max}} = 50 M_\odot$ but steeper power-law slope $\alpha = 2.35$ (left column). As the first four LIGO detections currently favor $M_{\text{max}} < 50 M_\odot$ and moderately shallow power-law slopes $\alpha < 3$ (Figure 4), we expect that fewer than 40 detections will strongly constrain $M_{\text{max}}$.

5. Discussion

5.1. Effect of Redshift Evolution

We have assumed that the merger rate, as measured in the source-frame, is uniform in comoving volume (Equation (11)). In reality, the merger rate per comoving volume is expected to increase with redshift until $z \sim 2$ (see, for example, Extended Data Figure 4 in Belczynski et al. 2016b). 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 $(VT)_\text{low}/(VT)_\text{high}$, displayed as a function of power-law slope $\alpha$ in Figure 2, 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.

5.2. Distribution of Mass Ratios

In fitting the mass distribution of BBHs (Equation (11)), we assumed that the distribution of mass ratios, $q$, is uniform in the allowed range $M_{\text{min}}/m_1 < q < \min(M_{\text{total, max}}/m_1 - 1, 1)$. In particular, we assumed that for a given primary component mass, $m_1$, the marginal distribution of $m_2 = qm_1$ is given by Equation (2). However, many BBH formation models predict a preference for equal-mass mergers (Dominik et al. 2012; Rodriguez et al. 2016). To explore the effects of our assumed mass ratio distribution, we generalize Equation (2):

$$p(m_2|m_1) \propto \frac{m_2^k}{\min(m_1, M_{\text{total, max}} - m_1)^{k+1} - M_{\text{min}}^{k+1}},$$

so that $k = 0$ reduces to Equation (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 $(\alpha, M_{\text{max}})$ if we fix $k = 6$ rather than $k = 0$. However, we note that there is currently no evidence that the distribution of mass ratios, $p(q|m_1)$, 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 $[M_{\text{min}}/m_1, 1]$. We can explicitly check if the data favor $k > 0$ if we incorporate Equation (17) into the power-law mass distribution, so that Equation (11) becomes

$$p(m_1, m_2|\alpha, M_{\text{max}}) \propto \frac{m_1^{-\alpha} m_2^k \mathcal{H}(M_{\text{max}} - m_1)}{\min(m_1, M_{\text{tot,max}} - m_1)^{k+1} - M_{\text{min}}^{k+1}}.$$  

We fit the above Equation (18) for $k$, marginalizing over $\alpha$ and $M_{\text{max}}$, and find that, for the first four LIGO detections, the likelihood $p(d|k)$ 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.

5.3. 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 mass (Belczynski et al. 2016b; Spera et al. 2016). 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 (O’Leary et al. 2016). 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, $M_{\text{max}}$) that fits for the bottom edge of the upper mass gap.

5.4. 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 Abbott et al. (2017e). 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 (Belczynski et al. 2016b). 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 $M_{\text{ZAMS}} < 150 M_\odot$ (Belczynski et al. 2016a; Eldridge & Stanway 2016; Spera et al. 2016). However, stars with $M_{\text{ZAMS}} \gtrsim 200 M_\odot$ in a sufficiently low-metallicity environment...
through NSF grant PHY-1125897 and an endowment from the Kavli Foundation. D.E.H. 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.

ORCID iDs
Maya Fishbach @ https://orcid.org/0000-0002-1980-5293
Daniel E. Holz @ https://orcid.org/0000-0002-0175-5064

References
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016a, PhRvX, 6, 041015
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016b, PhRvD, 93, 122003
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016c, LRR, 19, 1
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016d, ApJL, 833, L1
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017a, PhRvD, 95, 042003
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017b, PhRvL, 118, 221101
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017c, arXiv:1711.05758
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017d, PhRvL, 119, 141101
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017e, arXiv:1704.06428
Ade, P. A. R., Aghanim, N., Arnaud, M., et al. 2016, A&A, 594, A13
Belczynski, K., Buonanno, A., Ciardullo, R., et al. 2014, ApJ, 795, 120
Belczynski, K., Heger, A., Gladsy, W., et al. 2016a, A&A, 594, A9
Belczynski, K., Holz, D. E., Bulik, T., & O’Shaughnessy, R. 2016b, Natur, 534, 512
Belczynski, K., Repetto, S., Holz, D. E., et al. 2016c, ApJ, 819, 108
Belczynski, K., Wiktorowicz, G., Fryer, C. L., Holz, D. E., & Kalogera, V. 2012, ApJ, 757, 91
Bond, J. R., Arnett, W. D., & Carr, B. J. 1984, ApJ, 280, 825
Caldéron Bustillo, J., Laguna, F., & Shoemaker, D. 2017a, PhRvD, 95, 104038
Caldéron Bustillo, J., Salemi, F., Dal Canton, T., & Jani, K. 2017b, arXiv:1711.02009
Cannon, K., Cariou, R., Chapman, A., et al. 2012, ApJ, 748, 136
Capano, C., Pan, Y., & Buonanno, A. 2014, PhRvD, 89, 102003
Dal Canton, T., & Harfst, S. 2016, PhRvD, 94, 023001
Dominik, M., Belczynski, K., Fryer, C., et al. 2012, ApJL, 759, 52
Dominik, M., Merritt, E., & O’Shaughnessy, R. 2015, ApJ, 806, 263
Eldridge, J. J., & Stanway, E. R. 2016, MNRAS, 462, 3302
Farr, W. M., Sivan, N., Cantrell, A., et al. 2011, ApJ, 741, 103
Finn, L. S., & Chernoff, D. F. 1993, PhRvD, 47, 2198
Fishbach, M., Holz, D. E., & Farr, B. 2017, ApJL, 840, L24
Fowler, W. A., & Hoyle, F. 1964, ApJ, 139, 9
Gerosa, D., & Berti, E. 2017, PhRvD, 95, 124046
Heger, A., & Woosley, S. E. 2002, ApJ, 567, 532
Kass, R. E., & Raftery, A. E. 1995, J. Am. Stat. Assoc., 90, 773
Klimenko, S., Yakuish, I., Mercer, A., & Mitselmakher, G. 2008, CQGra, 25, 114029
Kovetz, E. D., Cholis, I., Breyssse, P. C., & Kamionkowski, M. 2017, PhRvD, 95, 103010
Kreidberg, L., Builyn, C. D., Farr, W. M., & Kalogera, V. 2012, ApJ, 757, 36
Kroupa, P. 2001, MNRAS, 322, 231
Littenberg, T. B., Farr, B., Couglin, S., Holz, D. E., & Rosado, D. E. 2015, ApJL, 807, L24
Mandel, I., Farr, W., & Gair, J. 2016, Tech. Rep.P1600187, LIGO
Mandel, I., Farr, W. M., Colonna, A., et al. 2017, MNRAS, 465, 3254
Marchant, P., Langer, N., Podsiajloowski, P., Tauris, T. M., & Moriya, T. J. 2016, A&A, 588, A50
Messick, C., Blackburn, K., Brady, P., et al. 2017, PhRvD, 95, 042001
Nitz, A. H., Dent, T., Dal Canton, T., Fairhurst, S., & Brown, D. A. 2017, ApJ, 849, 118
O’Leary, R. M., Meiron, Y., & Kocsis, B. 2016, ApJL, 824, L12
 Özel, F., Psaltis, D., Narayan, R., & McClintock, J. E. 2010, ApJ, 725, 1918
Rakavy, G., Shaviv, G., & Zinamon, Z. 1967, ApJ, 150, 131
Rodriguez, C. L., Chatterjee, S., & Rasio, F. A. 2016, PhRvD, 93, 084029
Salpeter, E. E. 1955, ApJ, 121, 161
Slutskey, J., Blackburn, L., Brown, D. A., et al. 2010, CQGra, 27, 16023
Spera, M., Giacobbo, N., & Mapelli, M. 2016, MmSAI, 87, 575
Spera, M., & Mapelli, M. 2017, MNRAS, 470, 4739
Usman, S. A., Nitz, A. H., Harry, I. W., et al. 2016, CQGra, 33, 215004
Veitch, J., Raymond, V., Farr, B., et al. 2015, PhRvD, 91, 042003
Woosley, S. E. 2017, ApJ, 836, 244

We thank Will Farr for valuable discussions. M.F. was supported by the NSF Graduate Research Fellowship Program under grant No. 1746045. M.F. and D.E.H. 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.