Are merging black holes born from stellar collapse or previous mergers?

Davide Gerosa and Emanuele Berti

Phys. Rev. D **95**, 124046 — Published 26 June 2017

DOI: 10.1103/PhysRevD.95.124046

Are merging black holes born from stellar collapse or previous mergers?

Davide Gerosa[1, *] and Emanuele Berti[2, 3, †]

[1] *TAPIR 350-17, California Institute of Technology, 1200 E California Boulevard, Pasadena, CA 91125, USA*
[2] *Department of Physics and Astronomy, The University of Mississippi, University, MS 38677, USA*
[3] *CENTRA, Departamento de Física, Instituto Superior Técnico, Universidade de Lisboa, Avenida Rovisco Pais 1, 1049 Lisboa, Portugal*

(Dated: May 22, 2017)

Advanced LIGO detectors at Hanford and Livingston made two confirmed and one marginal detection of binary black holes during their first observing run. The first event, GW150914, was from the merger of two black holes much heavier that those whose masses have been estimated so far, indicating a formation scenario that might differ from "ordinary" stellar evolution. One possibility is that these heavy black holes resulted from a previous merger. When the progenitors of a black hole binary merger result from previous mergers, they should (on average) merge later, be more massive, and have spin magnitudes clustered around a dimensionless spin ~ 0.7. Here we ask the following question: can gravitational-wave observations determine whether merging black holes were born from the collapse of massive stars ("first generation"), rather than being the end product of earlier mergers ("second generation")? We construct simple, observationally motivated populations of black hole binaries and we use Bayesian model selection to show that measurements of the masses, luminosity distance (or redshift) and "effective spin" of black hole binaries can indeed distinguish between these different formation scenarios.

I. INTRODUCTION

The observation of gravitational waves (GWs) from merging black hole (BH) binaries was a milestone in physics and astronomy [1–3]. During their first observing run (O1), the Advanced LIGO detectors detected two GW events (GW150914 and GW151226) and a marginal candidate LVT151012, which is also likely to be of astrophysical origin. The second observing run (O2) is currently ongoing, and Advanced Virgo is expected to join the detector network soon. Dozens of BH mergers may be detected by the end of O2 or in the third run (O3), allowing for statistical studies of their populations.

These events can further our understanding of the formation channels of binary BHs [4], because different astrophysical scenarios predict different binary properties. As the number of detections grows, a statistical analysis of the observed binary parameters should eventually allow us to identify or constrain the physical processes responsible for the formation and merger of compact binaries. Currently favored scenarios include stellar evolution of field binaries [5] and the dynamical capture of BHs in globular clusters [6]. Recent work showed that both field formation [7–14] and cluster formation [15–18] are broadly compatible with current Advanced LIGO observations [4].

It is quite likely that both field and cluster formation channels are at work in nature. The first event, GW150914, was the most surprising, because the merging BHs are much heavier that those whose masses have been estimated so far in X-ray binaries [19, 20], indicating a formation scenario that might differ from "ordinary" stellar evolution. Alternative theoretical scenarios which could explain the unexpected properties of GW150914 include formation via hierarchical triples [21–23], a Population III origin for the binary members [24–26], chemically homogeneous evolution in short-period binaries [27–29], and a primordial origin for the merging BHs [30, 31].

One possibility to explain the high mass of the merging BHs in GW150914 is that these BHs did not form following stellar collapse, but rather from previous BH mergers. Field formation scenarios typically predict long delay times between the formation and merger of a BH binary [9], so repeated mergers seem unlikely. However gravitational encounters are more common in dense stellar environments, and some scenarios suggest that repeated mergers may be possible [32–34]. The most likely environment to host multiple mergers are nuclear clusters [32], which present larger escape speeds compared to globular and open clusters, and can therefore more easily retains merger remnants with substantial recoils [35]. Stellar-mass BH binaries may also form in AGN gaseous discs [36], where migration traps can be invoked to assemble multiple generations of mergers [37]. Primordial BHs are also expected to merge very quickly [30, 31], so the possibility of repeated mergers in this scenario should not be excluded [38].

In this paper we ask the following question: can GW observations determine whether merging BHs such as those in GW150914 were born directly from the collapse of massive stars ("first-generation" BHs, henceforth 1g) rather than being the end product of previous mergers ("second-generation" BHs, henceforth 2g)?

Roughly speaking, one can expect mergers to leave several statistically observable imprints in 2g BHs, namely:

* Einstein Fellow; dgerosa@caltech.edu
† eberti@olemiss.edu

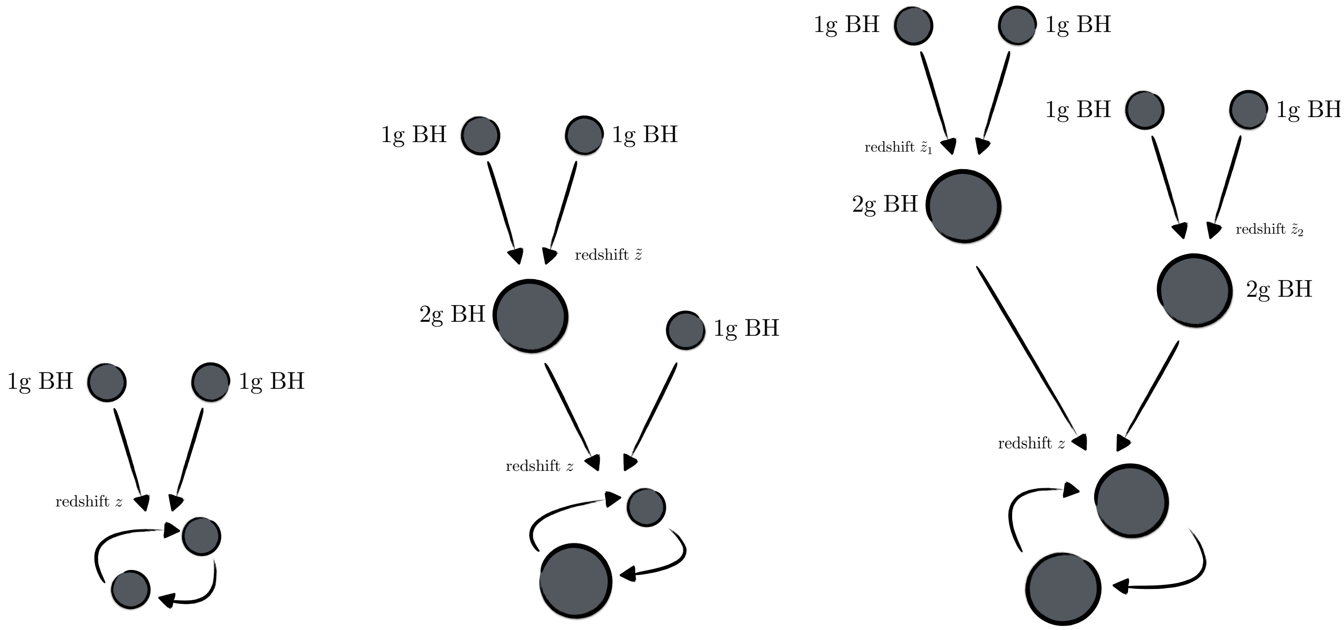

FIG. 1. Cartoon sketch of the three possible scenarios for the merger of two BHs. First generation (1g) BHs resulting from stellar collapse can form second generation (2g) BHs via mergers. Imprints of these formation channels are left in the statistical distribution of masses, spins and redshift of the detected events.

- 2g BHs should be more massive than BHs born from stellar collapse;
- quite independently of the distribution of spin magnitudes following core collapse (which is highly uncertain [39]), the spin magnitudes of 2g BHs should cluster (on average) around the dimensionless spin ~ 0.7 resulting from the merger of nonspinning BHs [40];
- statistically, the merger of BH binaries including 2g components should occur later (i.e., at smaller redshift or luminosity distance from GW detectors) because of the delay time between BH formation and merger.

In this paper we make these arguments more quantitative and rigorous by developing a simple but physically motivated model to describe the bulk theoretical properties of 1g and 2g binary BH mergers (Section II). Then we consider a set of present and future GW detectors and we simulate *observable distributions* by selecting detectable binaries and estimating the expected measurement errors on their parameters (Section III). Finally we set up a Bayesian model selection framework (Section IV) to address what can be done with current observations, and to quantify the capabilities of future detectors to distinguish between different models (Section V). We conclude by summarizing our results and pointing out possible extensions (Section VI).

II. THEORETICAL DISTRIBUTIONS

Our goal in this section is to develop a simple prescription to build populations of binary BHs. Our greatly oversimplified model is not meant to capture the complexity of binary evolution in an astrophysical setting, but just the main features distinguishing 1g and 2g BHs.

As illustrated by the cartoon in Figure 1, we construct three *theoretical distributions*, labeled by "1g+1g", "1g+2g" and "2g+2g". In this context, "1g" means that one of the binary components is a first-generation BH produced by stellar collapse, whereas "2g" means that it is a second-generation BH produced by a previous merger.

A. The 1g+1g population

Following the LIGO-Virgo Scientific Collaboration [3], for the 1g+1g population we adopt three different prescriptions for the distribution of source-frame masses:

(i) **Model "flat":** we assume uniformly distributed source-frame masses m_1 and m_2 in the range $m_i \in [5M_\odot, 50M_\odot]$ ($i = 1, 2$), where hereafter $m_1 > m_2$.

(ii) **Model "log":** we take the logarithm of the source-frame masses to be uniformly distributed in the same range, so that the probability distribution $p(m_1, m_2) \propto 1/m_1 m_2$.

(iii) **Model "power law":** we adopt a power-law distribution with spectral index $\alpha = -2.5$ for the primary BH (i.e. $p(m_1) \propto m^{\alpha}$), while the secondary mass is uniformly distributed in $m_2 \in [5M_\odot, m_1]$.

The upper limit of $50M_\odot$ was chosen to be consistent with current LIGO compact binary coalescence searches, and it excludes "by construction" intermediate-mass BH searches, discussed e.g. in [41]. Moreover, pair instability and pulsation pair instability in massive helium cores [42, 43] may inhibit the formation of 1g BHs with masses larger than $\sim 50M_\odot$ [44]. If multiple mergers occur through mass segregation in stellar clusters, the more massive objects will tend to form binaries, thus increasing the component masses of 1g+2g and 2g+2g populations. Our $50M_\odot$ upper mass limit is therefore conservative, because physical mechanisms such as pair instabilities and mass segregation would further separate the mass distributions of populations involving multiple mergers and make them more easily distinguishable.

Given the great uncertainties on the spin magnitude and orientation of binary BHs [45–48], in all three cases we assume the dimensionless spin magnitudes $\chi_{1,2}$ to be uniformly distributed in $[0, 1]$, and their directions to be isotropically distributed[1]. We are only interested in the global statistical properties of the population. Since isotropic spin distributions stay isotropic under precession and gravitational radiation reaction [51, 52], the assumption of isotropy will hold also at the small separations relevant for GW observations. For this reason there is no need to carry out post-Newtonian evolutions of the spin distributions for individual binaries of the kind discussed in [52–54].

B. The 2g+2g population

In order to construct the 2g+2g population we use the following procedure. We randomly extract two binaries from a given 1g+1g population. For these binaries, we estimate the final mass M_f and spin $\chi_{\rm f}$ of the merger remnant using the numerical relativity fitting formulas of Refs. [55, 56][2] as implemented in [54]. These masses and spins are used as input for the second round of binary mergers.

To perform meaningful comparison with the 1g+1g model described above, we again restrict our population to binaries with component masses in the range $[5, 50]M_\odot$, because this is the mass range targeted by LIGO compact binary coalescence searches.

C. The 1g+2g population

The 1g+2g distribution is the obvious mixture of the two: we draw one binary from the 1g+1g distribution, merge it to obtain a 2g BH, and then consider the merger of this 2g BH with a 1g BH.

D. Redshift distribution

The redshift distribution of BH mergers in the three different populations should be different, because on average 2g mergers are expected to happen later than 1g mergers. We can estimate the delay times between the formation and merger of a BH binary using the quadrupole formula

$$\frac{da}{dt} = -\frac{64}{5}\frac{q}{(1+q)^2}\frac{M^3}{a^3}\frac{G^3}{c^5}\,, \tag{1}$$

with the result

$$t = \int_a^0 \frac{dt}{da'}da' = \frac{5}{256}\frac{(1+q)^2}{q}\frac{a^4}{M^3}\frac{c^5}{G^3}\,. \tag{2}$$

If the binary initial separations a are drawn from a log-flat distribution (i.e., $dn/da \propto 1/a$), the distribution of the merger times is also log-flat (cf. [26]):

$$\frac{dn}{dt} = \frac{dn}{da}\frac{da}{dt} \propto \frac{1}{a^4} \propto \frac{1}{t}\,. \tag{3}$$

The "lookback time" t_L is given by [64]

$$t_L = \frac{1}{H_0}\int_0^z \frac{dz}{(1+z)\sqrt{\Omega_M(1+z)^3 + \Omega_\Lambda}}\,, \tag{4}$$

where we assume $\Omega_k = 0$, $\Omega_M = 0.307$, $\Omega_\Lambda = 0.693$ and $H_0 = 67.7\,\mathrm{km s^{-1} Mpc^{-1}}$ [65]. From the lookback time we can compute the time $t_L(z_1) - t_L(z_2)$ necessary for the Universe to evolve from redshift z_1 to redshift z_2.

We distribute the 1g+1g sources uniformly in comoving volume with redshifts $z < 2$. For the 1g+2g population, we assume that 2g BHs formed as some redshift $\tilde{z}$ drawn from the same distribution used for 1g+1g binaries. We then extract a delay time t_D from a flat distribution in $\log(t_D)$ in the range $t_D \in [10^{-4}\ \mathrm{Gyrs}, t_L(\tilde{z})]$. The lower limit is very conservative, and it roughly corresponds to the merger time for a $10M_\odot$ BH binary evolving from an initial orbital separation $a = 10R_\odot$. The

[1] Rodriguez et al. [48] argued that massive field binaries should typically have aligned spins because "heavy" BHs receive small supernova kicks that are unable to tilt the orbit [49, 50], while the spins of massive binaries produced in dense stellar environments should be isotropically distributed. A more detailed investigation of the correlation between spin alignment and binary BH formation requires astrophysical modeling that is beyond the scope of this paper (see e.g. [45, 50]).

[2] There are several alternative fitting formulas for the final masses and spins [57–63]. The difference between different prescriptions is smaller than measurement errors in GW observations, and therefore the choice of a particular fitting formula is of no consequence for our present purpose.

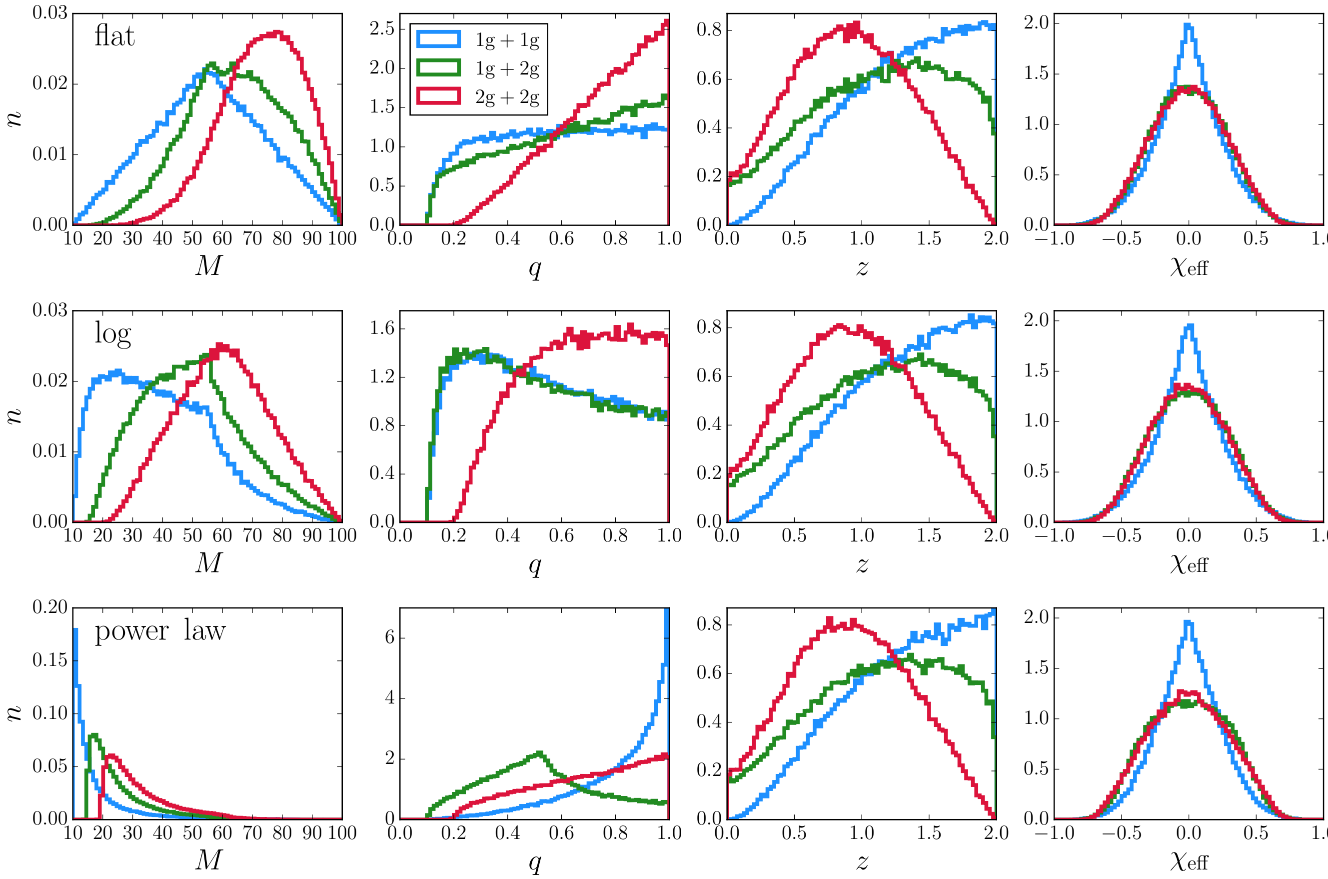

FIG. 2. Theoretical distribution of the observable parameters $\mathbf{u} = \{M,\, q,\, z,\, \chi_{\rm eff}\}$ for 1g+1g (blue), 1g+2g (green) and 2g+2g (red) populations, assuming the "flat" (top), "log" (middle) and "power law" (bottom) mass distributions.

redshift z of a 1g+2g merger is then given by the numerical solution of the equation

$$t_L(\tilde{z}) - t_L(z) = t_D\,. \tag{5}$$

Finally, for the 2g+2g population we extract two values $\tilde{z}_1$, $\tilde{z}_2$ from the 1g+1g distribution. The redshift z of a 2g+2g merger follows again from a numerical solution of Eq. (5), with the difference that now we set $\tilde{z} = \min(\tilde{z}_1, \tilde{z}_2)$.

In Sec. V D we will discuss how time delay prescriptions affect our results.

E. Measurable parameters

For concreteness and simplicity, we will characterize each binary by its total mass $M = m_1 + m_2$, mass ratio $q = m_2/m_1 \leq 1$, redshift z and "effective spin" [66]

$$\chi_{\rm eff} = \frac{1}{M}\left(\frac{\mathbf{S_1}}{m_1} + \frac{\mathbf{S_2}}{m_2}\right)\cdot \hat{\mathbf{L}}\,. \tag{6}$$

The effective spin (a mass-weighted sum of the projection of the spins $\mathbf{S_i} = m_i^2 \chi_i \hat{\mathbf{S}}_\mathbf{i}$ along the orbital angular momentum $\mathbf{L}$) is a constant of the motion in post-Newtonian evolutions, at least at 2PN order [52, 67]. It is also the easiest spin parameter to measure [66, 68].

Let us introduce a vector $\mathbf{u}$ whose components are the observable variables to use in our statistical analysis, i.e.

$$\mathbf{u} = \{M,\, q,\, z,\, \chi_{\rm eff}\}\,. \tag{7}$$

The components of this vector will be labeled by an index $j = 1, ..., J$ such that $u_1 = M$, $u_2 = q$, etcetera; a capital Latin index J will denote the dimensionality of the vector $\mathbf{u}$, i.e. the number of observables considered in the analysis. Each binary in our catalog is characterized by a specific set of observable properties $\bar{u}^{(i)}$, where the superscript index ($i = 1, ..., I$) labels entries in our synthetic catalog.

The theoretical distributions of measurable source parameters $\mathbf{u} = \{M,\, q,\, z,\, \chi_{\rm eff}\}$ for 1g+1g, 1g+2g and 2g+2g events are compared in Figure 2. Each row corresponds to one of the three mass distributions described in Section II A.

The mass distributions have some noteworthy features. First of all, and quite obviously, 2g BHs have higher component masses. Therefore the total mass is higher when 2g BHs are present (for any given assumption on the mass distribution), and this effect is most notable

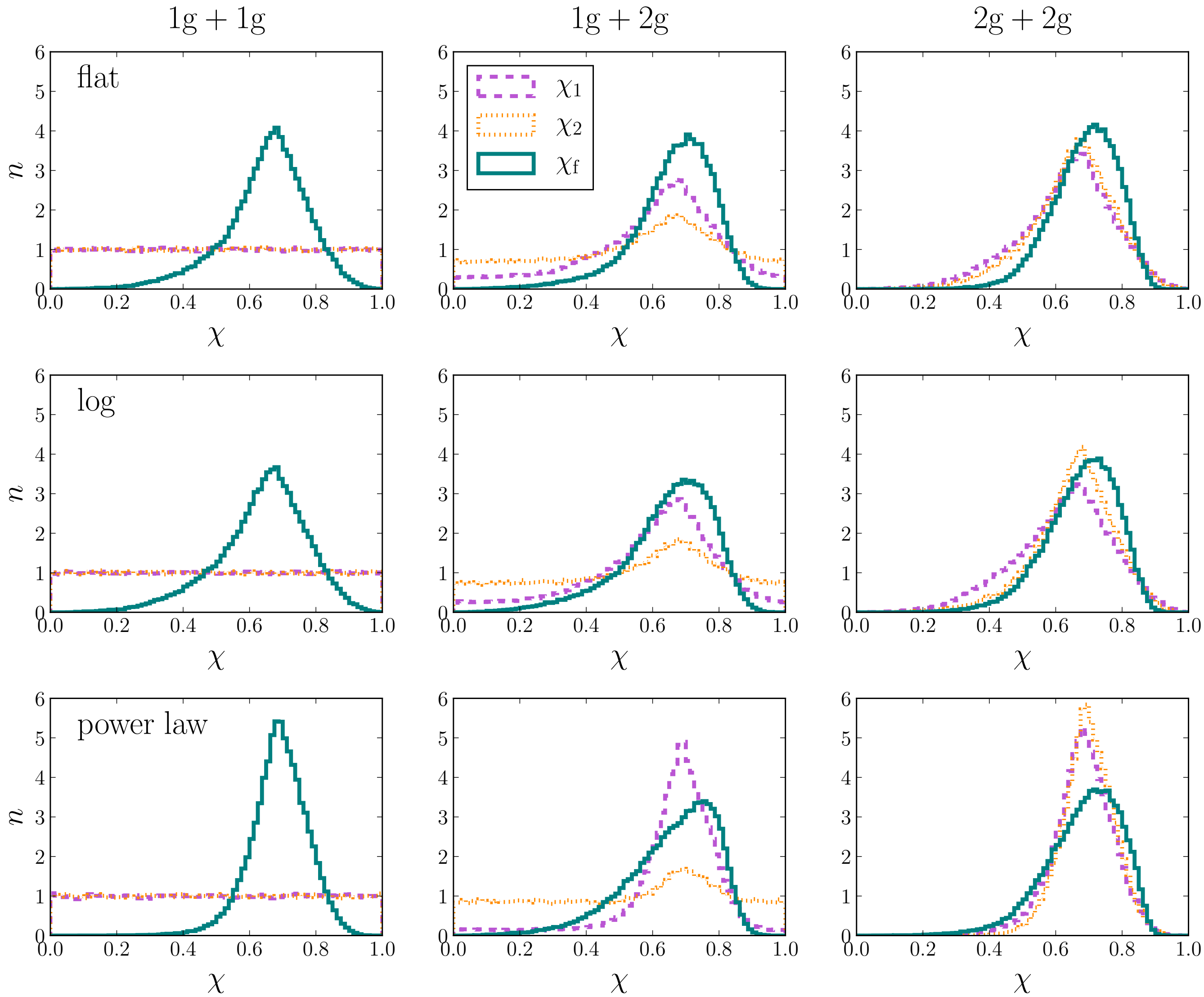

FIG. 3. Spin magnitude distributions for primary (χ_1), secondary (χ_2) and post-merger (χ_f) BH spins in each of the various models used in this paper. On average, mergers tend to produce BH spins clustered around ~ 0.7, quite independently of the progenitor parameters (cf. Figure 3 and the left panels in Figures 4 and 5 of Ref. [40]).

for the 2g+2g distributions. Mergers also tend to increase the number of comparable-mass binaries, in part because of the fixed mass range for the component masses ($m_i \in [5, 50]M_\odot$). For the "power law" mass function, the mass ratio of the 1g+2g population peaks at $q = 0.5$. This is because the mass distribution of the primary BH is strongly peaked at the low end of the range (i.e., at $\sim 5M_\odot$), so many 2g binaries are nearly equal mass, with component masses close to $5M_\odot$.

Redshift distributions also follow the expected trend: most 1g+1g events occur at large redshift, whereas mergers involving one or two 2g BHs occur (on average) at smaller redshift, because there is a time delay between the formation of 1g BHs via core collapse and their subsequent merger.

The most striking differences are found in the distributions of individual spins. To better illustrate this point, in Figure 3 we show the distribution of the individual BH spins (χ_1, χ_2), as well as the distribution of the spin of the remnant χ_f. As discussed in [40], from a statistical point of view the effect of mergers is to "cluster" BH spins around $\chi_f \sim 0.7$, quite independently of the progenitor parameters. While the 1g+1g spin magnitudes are uniform in the range [0, 1] by construction, spin distributions become peaked at ~ 0.7 when 2g BHs are involved. This clustering is evident in the distribution of primary spins χ_1 for the 1g+2g and 2g+2g cases, and in the distribution of secondary spins χ_2 for the 2g+2g case. For the 1g+2g population, the peak at $\chi_2 \sim 0.7$ is less pronounced. This is because the lower-mass BH is most likely 1g, and the spin distribution of 1g BHs is by construction uniform in [0, 1].

Unfortunately low-SNR GW observations of merger events are not very sensitive to χ_1 and χ_2, but rather to the effective spin $\chi_{\rm eff}$ defined in Eq. (6). The right column of Figure 2 shows that the effect of mergers is considerably smeared out in $\chi_{\rm eff}$, but more binaries with $\chi_{\rm eff} \sim 0$ are expected if all sources are 1g BHs. Measurements of $\chi_{\rm eff}$ may still be sufficient to distinguish between different populations, especially when comparing 1g+1g against either 1g+2g or 2g+2g. Discriminating between BH progenitors should be considerably easier with future detectors, when high-SNR events will allow for more precise measurements of χ_1, χ_2 and $\chi_{\rm f}$ [69–71] .

F. Single detections

In the rest of this paper we will study how statistical inference from several detections can be used to constrain the underlying BH population. However, it is possible that *single* detections with specific parameters can already provide smoking gun evidence for the occurrence of multiple mergers.

One possibility, as mentioned in Section II A, is that pair instabilities may prevent the formation of 1g BHs with masses sensibly above $\sim 50 M_\odot$ [42–44]. If this is indeed the case, a single detection of a merging BH binary where one of the components has mass larger than $50 M_\odot$ would indicate the occurrence of multiple mergers. This argument, however, relies on two crucial assumptions: (i) that 1g BHs always form from stellar collapse, while more exotic formation channels (e.g. involving primordial BHs) may produce massive BHs without invoking multiple mergers; (ii) that pair instabilities in core collapse do indeed prevent the formation of massive BHs. Pair instabilities, pair instability pulsations and the exact value of the maximum BH mass that can be produced via core collapse are all topics of current research [44].

Another possibility involves accurate measurements of the component spins through the detection of a single nearby, non face-on binary merger with comparable, low masses and many precession cycles in the LIGO band. Unfortunately, parameter estimation studies suggest that current-generation detectors could allow dimensionless spin measurement errors ~ 0.3 in best-case scenarios [72]. Errors of this magnitude are comparable to the width of the peaks in the spins distributions shown in Fig. 3 and there is significant uncertainty in the spin magnitude distribution of astrophysical BHs, so it seems unlikely that *single* spin measurements may allow us to tell apart 1g BHs from 2g BHs, at least in the near future.

III. OBSERVABLE DISTRIBUTIONS

From the *theoretical distributions* described in Section II, we construct *observable distributions* by (i) selecting detectable binaries according to a detection statistic, such as a threshold in the signal-to-noise ratio (SNR), and (ii) folding in measurement errors.

A. Detection probability

We first assign a detection probability $\kappa^{(i)} < 1$ to each binary in our catalogs. This number takes into account the detector sensitivity and antenna pattern, as well as the (random) sky position of the source. We compute $\kappa^{(i)}$ following the procedure outlined in Ref. [12], where an astrophysical catalog of binaries produced using the StarTrack population synthesis code was filtered to produce similar catalogs of observable binaries for a specific set of GW detectors. This procedure is briefly reviewed below.

Each binary produces a GW strain $h(t)$ and an expectation value for the SNR

$$\rho^2 = 4 \int_0^\infty \frac{|\tilde{h}(f)|}{S_n(f)} df \,, \tag{8}$$

where $S_n(f)$ is the noise power spectral density of the detector and $\tilde{h}(f)$ is the Fourier transform of the strain $h(t)$. The strain is computed using the IMRPhenomC waveform model [73]. In this paper we consider noise power spectral densities for the first AdLIGO observing run (O1), the Advanced LIGO design sensitivity [74], A+ (Advanced LIGO with squeezing) and Voyager (the most advanced instrument that can be hosted in facilities similar to LIGO) [75].

For any binary in our catalog we can compute $\rho_{\rm opt}$, i.e. the single-detector SNR for a binary that is optimally located and oriented in the sky. We then select those binaries in the catalog that are above a detection threshold $\rho_{\rm opt} \geq \rho_{\rm thr} = 8$. This criterion has often been used as a simple, reasonable proxy for a more realistic calculation of GW detection rates in multi-detector networks [12, 76]. Then we compute the detection probability as

$$\kappa^{(i)} = P\left(w^{(i)}\right) \,, \tag{9}$$

where the function $P(w^{(i)})$ is the cumulative distribution function for the projection parameter $w^{(i)} \equiv \rho_{\rm thr}/\rho^{(i)}_{\rm opt}$. This cumulative distribution function takes into account the geometrical "peanut factor" that characterizes the sensitivity of the detector to the source sky location, inclination and polarization (see [12] and references therein). Roughly speaking, $w^{(i)} = 1$ means that the source is in a "blind spot" of the detector, while $w^{(i)} = 0$ in the high-SNR limit. A tabulated version of $P(w^{(i)})$ is publicly available[3]; we use standard spline interpolation to compute this function for generic values of $w^{(i)}$.

[3] www.phy.olemiss.edu/~berti/research

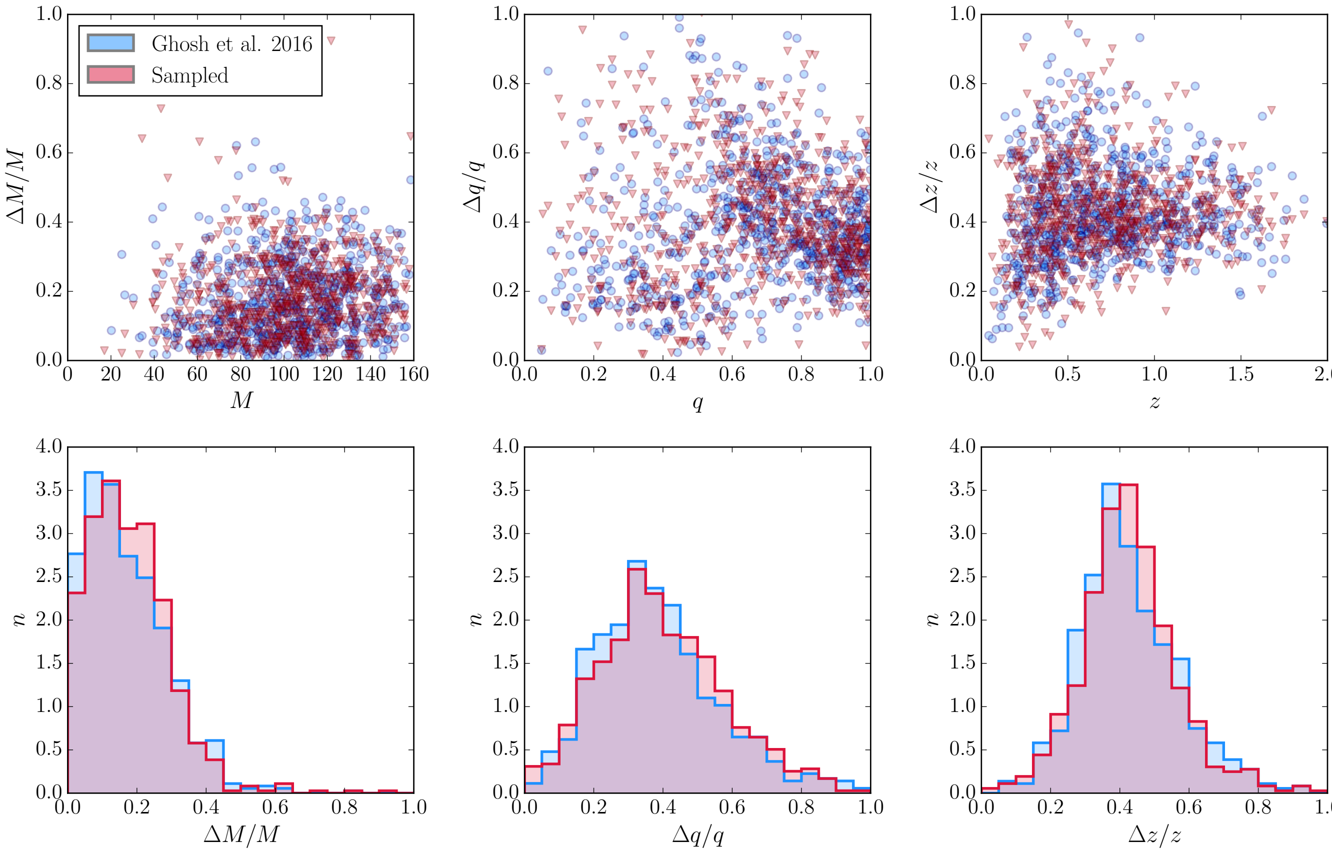

FIG. 4. Blue: relative errors on the total mass M (left), mass ratio q (middle) and redshift z (right) as computed in Ref. [77]. Red: resampling of these data, obtained as described in Section III B. The top panels show scatter plots of the relative error on each parameter as a function of the value of that parameter for the source. The bottom panels show the same information as a histogram.

B. Measurement errors

Ideally we should compute measurement errors for each binary in the catalog using Markov-Chain Monte Carlo methods and use the obtained posteriors to perform model selection. This is computationally expensive, and unnecessary from the point of view of our proof-of-principle analysis. For our present purpose we adopt a much simpler prescription, described below.

We build on study by Ghosh *et al.* [77], who computed BH binary measurement errors using the LALINFERENCE code [78] (see also [79–82] for more work on the subject). In particular, we use their results for aligned-spin BH binaries detected by a network of 3 advanced detectors. Their data set provides 1σ errors on several quantities, including the total mass M, mass ratio q and redshift z. These are shown in blue in Figure 4.

The data set is too sparse to perform an efficient binning and interpolation in three dimensions (M, q, z). In order to partially account for the expected degeneracies (e.g., close binaries will generally have smaller errors on the masses), we adopt the following procedure. Consider a binary in our catalog with parameters $\bar{M}$, $\bar{q}$, $\bar{z}$. To estimate measurement errors on the parameters of this binary, we consider the 5 "closest" binaries in the data set of Ref. [77], and compute the average and standard deviation of their measurement errors. Here "closest" is defined in the following sense: given the maximum and minimum value of each of the three parameters (M, q, z), we rescale their actual values so that these parameters are distributed in a cube of size one; then we compute the Euclidean distance between binaries in this cube. The average and standard deviation from the 5 closest binaries are then used to extract the measurement errors $\sigma_{\bar{M}}, \sigma_{\bar{q}}, \sigma_{\bar{z}}$ from a normal distribution. The red dots and histograms in Figure 4 show the measurement errors obtained from this resampling. The obtained distributions look remarkably close to the original data. Errors on the redshift are slightly overestimated, so (if anything) our resampling procedure seems to yield conservative predictions. Estimates for the errors on $\chi_{\rm eff}$ were not computed in Ref. [77], so we assume $\sigma_{\chi_{\rm eff}} = 0.1$ for all binaries measured by LIGO at design sensitivity. This rough estimate is quite conservative, and it is consistent with measurement errors in the first GW detections [3].

Ref. [77] computed parameter estimation errors for the

LIGO-Virgo network at design sensitivity. Fisher matrix arguments [83] suggest that the capabilities of other detectors can be estimated rescaling the errors on the total mass, mass ratio, luminosity distance and $\chi_{\rm eff}$ by the ratio of SNRs, i.e.

$$\sigma_{\rm Detector} = \sigma_{\rm LIGO} \frac{\rho_{\rm LIGO}}{\rho_{\rm Detector}} \,. \tag{10}$$

Luminosity distance and redshift are related by

$$D_L = \frac{1+z}{H_0} \int_0^z \frac{dz}{\sqrt{\Omega_M (1+z)^3 + \Omega_\Lambda}} \,, \tag{11}$$

(where we use units such that $c = 1$), so that

$$\frac{dD_L}{dz} = \frac{D_L}{1+z} + \frac{1+z}{H_0 \sqrt{\Omega_M (1+z)^3 + \Omega_\Lambda}} \,. \tag{12}$$

The error on the redshift σ_z is related to the error on D_L by

$$\left(\frac{\sigma_z}{D_L} \frac{dD_L}{dz} \right)^2 = \left(\frac{\sigma_{D_L}}{D_L} \right)^2 + \left(\frac{\sigma_{H_0}}{H_0} \right)^2 \tag{13}$$

where we assumed $\sigma_{\Omega_\Lambda}, \sigma_{\Omega_\Lambda} \simeq 0$ (see e.g. [84, 85]). Given recent discrepancies in the determination of H_0, we assume $\sigma_{H_0}/H_0 = 0.1$ [86–88].

C. Binning

Recall that each binary is characterized by a vector of observable parameters $\mathbf{u} = \{u_1, \ldots, u_J\}$. If (for simplicity) we momentarily neglect measurement errors, the observable distribution is just a sum of Dirac deltas centered at $\bar{\mathbf{u}}^{(i)}$, and each delta is weighted by the detection probability $\kappa^{(i)}$:

$$\tilde{r}(\mathbf{u}, \lambda) = \frac{\sum_{i=1}^{I} \kappa^{(i)} \prod_{j=1}^{J} \delta \left(u_j - \bar{u}_j^{(i)} \right)}{\sum_{i=1}^{I} \kappa^{(i)}} \,, \tag{14}$$

where λ labels the model (cf. Section IV) and the denominator ensures normalization. Using the procedure described in Section III B we can obtain estimates of the 1σ errors on the measurement of each parameter. The i-th binary in the catalog now has estimated parameters $\bar{u}^{(i)}$ with errors $\sigma^{(i)} = \sigma(\bar{u}^{(i)})$. Assuming that errors are normally distributed and neglecting degeneracies, we can substitute the Dirac deltas of Eq. (14) with Gaussian distributions:

$$\tilde{r}(\mathbf{u}, \lambda) = \frac{\sum_{i=1}^{I} \kappa^{(i)} \prod_{j=1}^{J} \mathcal{N} \left(u_j; \bar{u}_j^{(i)}, \sigma^{(i)} \right)}{\sum_{i=1}^{I} \kappa^{(i)}} \,, \tag{15}$$

where

$$\mathcal{N} \left(u_j; \bar{u}_j^{(i)}, \sigma^{(i)} \right) = \frac{1}{\sigma^{(i)} \sqrt{2\pi}} \exp \left(- \frac{u_j - \bar{u}_j^{(i)}}{2\sigma^{2\,(i)}} \right) . \tag{16}$$

Next, we need to bin the distributions $\tilde{r}(\mathbf{u}, \lambda)$. In each direction j, we construct bins k_j with extrema b_{k_j} and B_{k_j}, i.e. $u_j \in (b_{k_j}, B_{k_j})$. The function $\tilde{r}(\mathbf{u}, \lambda)$ in each multi-dimensional bin $\{k_1, ..., k_J\}$ is given by the integral

$$\begin{aligned} \tilde{r}_{k_1,\ldots,k_J}(\lambda) &= \int_{b_{k_1}}^{B_{k_1}} du_1 \ldots \int_{b_{k_J}}^{B_{k_J}} du_J \; \tilde{r}(\mathbf{u}, \lambda) \\ &= \frac{\sum_{i=1}^{I} \kappa^{(i)} \prod_{j=1}^{J} \int_{b_{k_j}}^{B_{k_j}} \mathcal{N} \left(u_j; \bar{u}_j^{(i)}, \sigma^{(i)} \right) du_j}{\sum_{i=1}^{I} \kappa^{(i)}} \,. \end{aligned} \tag{17}$$

In practice, we spread each source over multiple bins because of measurement errors (see [89, 90] for a similar approach in the LISA context). Eq. (17) is correctly normalized to 1 only if the bins k_j span the entire support of $\tilde{r}(u, \lambda)$. When substituting Dirac deltas with Gaussian distributions we are adding support in the whole range $[-\infty, +\infty]$ for each of the u_j's, and inevitably we end up using a finite range. For simplicity, we just renormalize $\tilde{r}_{k_1,\ldots,k_J}(\lambda)$ such that

$$\sum_{k_1} \ldots \sum_{k_J} \tilde{r}_{k_1,\ldots,k_J} = 1 \,. \tag{18}$$

From now on we will identify the bins by a multi-index variable $k = \{k_1, ..., k_J\}$, so (for example) we can write $\sum_k f_k \equiv \sum_{k_1} \ldots \sum_{k_J} f_{k_1,\ldots,k_J}$ for any binned quantity f.

D. Putting the pieces together

Examples of observable distributions are given in Figure 5 for Advanced LIGO at design sensitivity (top) and Voyager (bottom) assuming the "flat" mass function. In each panel, dashed lines show the theoretical distribution for the 1g+1g, 1g+2g and 2g+2g populations, as already presented in Figure 2. The histograms show the *observable* population, i.e. the distribution of detectable binaries, where the measured parameters take into account also measurement errors. Some trends are visible.

Let us first focus on the top row, which refers to observations with Advanced LIGO at design sensitivity. It is clear that binaries with larger total mass and lower redshift produce stronger signals, and therefore they are more likely to be detected. In particular, Advanced LIGO can hardly detect any binaries at redshift $z \gtrsim 1$. The distribution of $\chi_{\rm eff}$ also shows a mild excess of observable events with $\chi_{\rm eff} \simeq 0$ for the 1g+1g population with respect to the 1g+2g and 2g+2g populations, suggesting that measurements of $\chi_{\rm eff}$ can indeed help to discriminate between populations.

The bottom row of Figure 5 shows that the increased sensitivity of a Voyager-like detector has two main effects: it makes observable distributions in each of the parameters much closer to the corresponding theoretical distributions, and (quite importantly) it extends the reach of the detector to high-z binaries. We obviously expect that more sensitive detectors will allow better discrimination between the different populations.

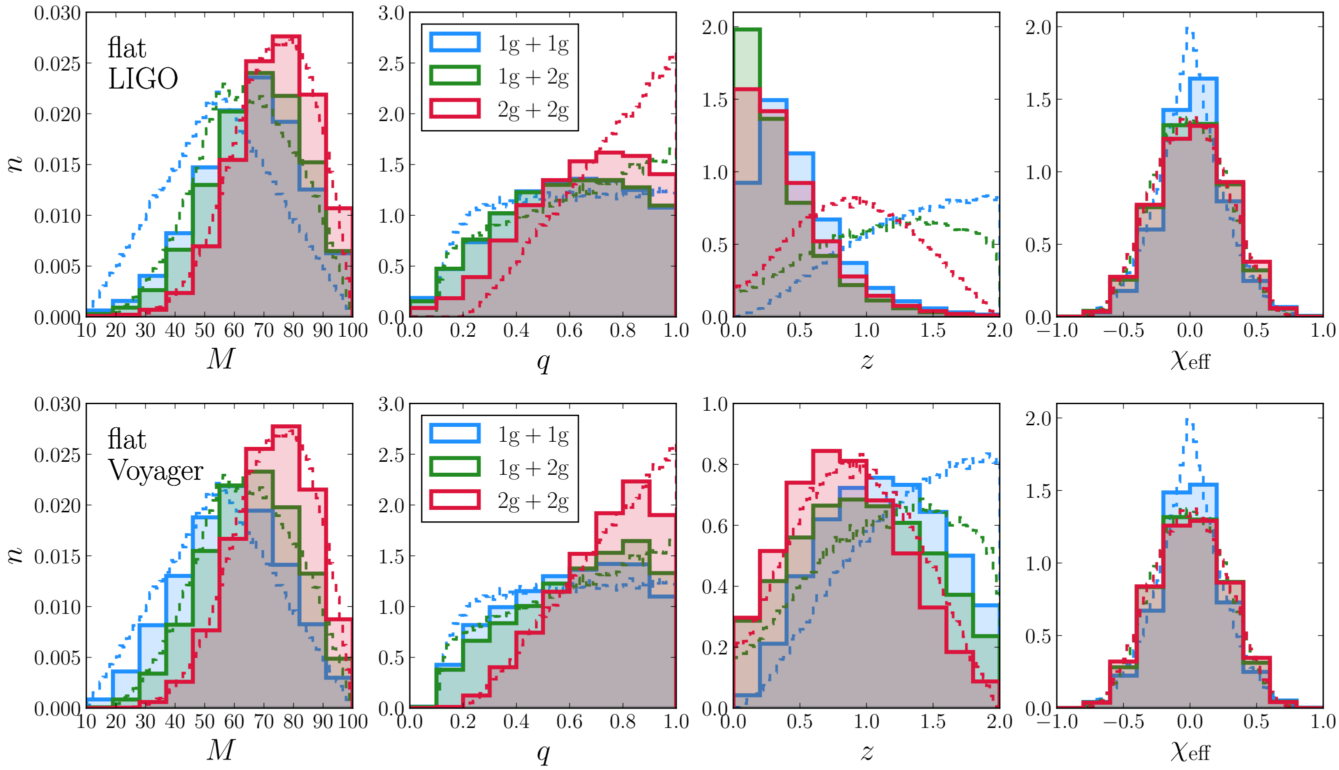

FIG. 5. Observable distributions for Advanced LIGO at design sensitivity (top) and Voyager (bottom). All plots refer to the "flat" mass distribution. In each panel, dashed lines show the theoretical distribution for the 1g+1g (blue), 1g+2g (green) and 2g+2g (red) populations; these are the same curves shown in Fig. 2. Following the same color scheme, solid shaded histograms show the "observed" population, consisting of events that pass the SNR threshold and that include measurement errors.

The 2g+2g population presents a peak at $M \sim 80M_\odot$ and $q \sim 1$. Equal mass binaries of $\sim 40M_\odot + 40M_\odot$ can only be detected by Advanced LIGO at design sensitivity if they are located at very small redshift (cf. e.g. [4]). This explains the significant drop in the number of observed events as $q \to 1$. The effect is strongly mitigated in Voyager, because the instrument is more sensitive at low frequency.

IV. STATISTICAL TOOLS

In this section we briefly introduce statistical tools to perform Bayesian model selection. We label models by a parameter λ that can be either discrete (if we want to distinguish two competing models A and B) or continuous (if want to measure the "mixing fraction" between competing models that best describes the data).

A. Number of observations

Our goal is to infer which model λ best describes a set of data. As explained above, our binned distributions $\tilde{r}_k(\lambda)$ are normalized. To compare our models with the data we need an extra parameter $N(\lambda)$, the total number of observations predicted by model λ. We write

$$r_k(\lambda) = N(\lambda)\, \tilde{r}_k(\lambda)\,. \tag{19}$$

As for the individual binary parameters, we introduce an array $\mathbf{d}$ whose elements are the single observations $d^{(i)}$, which in turn are $J-$dimensional arrays. We bin the array $\mathbf{d}$ on the same grid used for the catalogs to obtain binned values d_k.

The likelihood of obtaining a data set d_k from model λ is given by

$$p(\mathbf{d}|\lambda) = \prod_k \frac{(r_k(\lambda))^{d_k} e^{-r_k(\lambda)}}{d_k!}\,. \tag{20}$$

In our analysis the total number of observation does not contain information about the given model (this may not be the case for more realistic scenarios, where different models predict different merging rates: see e.g. [91]). We therefore marginalize the likelihood over $N(\lambda)$. Plugging Eq. (19) into Eq. (20) one obtains [90]

$$p(\mathbf{d}|\lambda) = \left(\prod_k \frac{(\tilde{r}_k(\lambda))^{d_k} e^{-\tilde{r}_k(\lambda)}}{d_k!}\right) \left(N(\lambda)^{\sum_k d_k} e^{-N(\lambda)}\right)\,, \tag{21}$$

		$1g{+}1g$ vs. $2g{+}2g$	$1g{+}1g$ vs. $1g{+}2g$	$1g{+}2g$ vs. $2g{+}2g$
O1 LIGO	flat	12.7 (15.8)	2.0 (2.0)	6.4 (7.6)
	log	3.3 (3.5)	0.9 (0.9)	3.5 (3.8)
	power law	0.7 (1.0)	1.3 (1.6)	0.6 (0.6)
Ad. LIGO (design)	flat	30.2 (37.8)	1.4 (3.7)	21.9 (10.11)
	log	4.3 (7.0)	0.6 (1.4)	6.9 (5.1)
	power law	0.6 (1.7)	1.0 (3.8)	0.6 (0.5)

TABLE I. Odds ratios from the three O1 observations (GW150914, GW151226 and LVT151012) and from hypothetical observations of the same events at Advanced LIGO design sensitivity. Odds ratios in parentheses were computed omitting all redshift information, i.e. considering the 3-dimensional vector of observables $\mathbf{u} = \{M,\, q,\, \chi_{\rm eff}\}$.

and consequently the marginalized likelihood is

$$\tilde{p}(\mathbf{d}|\lambda) = \left(\prod_k \frac{(\tilde{r}_k(\lambda))^{d_k} e^{-\tilde{r}_k(\lambda)}}{d_k!}\right) \sum_N \left(N^{\sum_k d_k} e^{-N}\right) . \tag{22}$$

Note that the term $\sum_N \left(N^{\sum_k n_k} e^{-N}\right)$ is a multiplicative coefficient that only depends on the data $\mathbf{d}$, and not on the model λ. This term can be ignored because, as we will see below, we are only interested in likelihood ratios, not in the likelihoods themselves.

From now on, to simplify notation, we will drop the tilde on p and assume that likelihoods are always marginalized over the total number of events.

B. Model selection

Let us first look at model comparison between *pure* models, so that λ is a discrete variable. Given models $\lambda = A$ and $\lambda = B$, their *odds ratio* is defined as

$$O_{AB} = \frac{p(\mathbf{d}|A)\pi(A)}{p(\mathbf{d}|B)\pi(B)} , \tag{23}$$

where π is the prior probability assigned to each of the two models. The simplest assumption on the priors is $\pi(A) = \pi(B) = 1/2$, such that the odds ratio reduces to the likelihood ratio. If $O_{AB} \gg 1$ ($O_{AB} \ll 1$) the data favors model A (B). The probability of model A is

$$p_A = \frac{O_{AB}}{1 + O_{AB}} = \frac{p(\mathbf{d}|A)}{p(\mathbf{d}|A) + p(\mathbf{d}|B)} , \tag{24}$$

and the probability of model B is $p_B = 1 - p_A$. Sometimes σ-levels are used to quantify the significance of a discrete model comparison, in analogy with Gaussian measurements. The expression relating the odds ratio $\mathcal{O}$ and σ is

$$\mathcal{O} = \frac{1}{1 - 2\,\mathrm{erf}(\sigma)} . \tag{25}$$

We can also assume that the data are represented by a mixture of two or more models, and assess whether the data themselves are informative about the underlying model mixing fractions. Each pure model m enters the mixed model with a weight f_m, such that $\sum f_m = 1$. Model comparison is equivalent to Bayesian inference on the parameters $\lambda = \{f_1, f_2, \ldots\}$, as described by the posterior distribution

$$p(\lambda|\mathbf{d}) = \frac{p(\mathbf{d}|\lambda)\pi(\lambda)}{\int p(\mathbf{d}|\lambda)\pi(\lambda) d\lambda} . \tag{26}$$

As before, $\pi(\lambda)$ is the prior assigned to each mixed model. We choose $\pi(\lambda)$ to be uniformly distributed on the surface $\sum f_m = 1$.[4] From a computational point of view, we first draw values of λ from the uniform prior, and then we produce a statistical sample distributed according to $p(\mathbf{d}|\lambda)$ using a standard Monte Carlo hit-or-miss algorithm.

V. RESULTS

So far we have outlined a procedure to build a set of "synthetic" GW observations of merging BH binaries (along with their associated errors) from simple astrophysical considerations. We now wish to understand whether these observations can be used to distinguish between different populations using Bayesian model selection (see e.g. [89–93] for previous studies of this problem in different contexts).

A. LIGO O1 data

We first apply our model comparison tool to the three LIGO O1 observations. The data set $\mathbf{d}$ consists of the maximum likelihood values provided in Ref. [3]:

- GW150914:
  $M = 65.3 M_\odot$, $q = 0.81$, $z = 0.090$, $\chi_{\rm eff} = -0.06$.

[4] For instance, for a mixture of three models $\lambda = \{f_1, f_2, f_3\}$ the equation $\sum f_m = 1$ describes a 2-dimensional surface S of area $\sqrt{3}/2$. The uniform prior on S is given by $\pi(f_1, f_2, f_3) = 2/\sqrt{3}$, so that $\iint_S \pi\, dS = 1$.

- GW151226:
  $M = 21.8 M_\odot$, $q = 0.52$, $z = 0.094$, $\chi_{\rm eff} = 0.21$.
- LVT151012:
  $M = 37 M_\odot$, $q = 0.57$, $z = 0.201$, $\chi_{\rm eff} = 0.03$.

As stressed above, measurements errors are included in this analysis at the level of the catalogs, by spreading each source over multiple bins. A more in-depth study should make use of the posterior distribution of the observed parameters obtained through dedicated parameter-estimation pipelines.

Performing model selection as described in the previous sections and using the O1 sensitivity curve, we obtain the odds ratio reported in Table I. We also repeat the same exercise assuming the anticipated noise power spectral density of Advanced LIGO at design sensitivity. This basically answers the question: "what if the O1 observations had been carried out with a better detector?"

As shown in Table I, most of the odds ratios are in the range $0.3 \lesssim \mathcal{O} \lesssim 3$, corresponding to 1σ. This simply indicates that 3 observations are not enough to perform a meaningful statistical analysis. However some of the comparisons return odds ratios $\mathcal{O} \sim 10$, approaching 2σ evidence. When this happens (i) the 1g+1g population seems to be preferred, and (ii) the odds become higher for a more sensitive detector like Voyager. In these cases the algorithm seems to capture real statistical differences between the catalogs, that become more pronounced when more binaries are detected and measurement errors get smaller.

As a note of caution, we stress here that such discrete model comparison analyses can only tell us which of two competing models better describes a given data set, *not* which model is correct. For instance, our results in Table I show some dependence on the underlying mass distribution. This could be due either to the low dimensionality of the statistical sample (cf. Section V B), or to the fact that none of the three mass distributions faithfully describes the observations. To bracket uncertainties in the time delay prescription (cf. Section II D), Table I also lists odds ratios computed omitting all redshift information. This calculation shows that assumptions on the delay times do not significantly affect our conclusions, given the limited statistics currently available. It will be straightforward to update our analysis with higher statistics and better motivated BH binary formation models when more data become available.

B. Simulated data: pure models

The results of Section V A show, not surprisingly, that more than 3 observations are needed to discriminate between different models. In order to estimate the capabilities of larger data sets and more sensitive detectors, here we perform model selection on simulated observations. Our main goal is to estimate how many observations are needed to distinguish a pair of models with a given confidence level.

Given a model $\lambda_{\rm true}$, we extract the number of events per bin d_k assuming a Poisson distribution

$$p(d_k) = \frac{r_k(\lambda_{\rm true})^{d_k} e^{-r_k(\lambda_{\rm true})}}{d_k!} \,. \tag{27}$$

Here the total number of observation $N_{\rm obs} = N(\lambda_{\rm true})$ is a free parameter that we need to specify. We expect model comparison to be easier/harder if more/less observations are available. This statement is made more quantitative in Figure 6 and Table II.

Figure 6 shows the odds ratio distribution obtained from several realization of $N_{\rm obs}$ observations. For each pair or models we plot $\mathcal{O}_{AB}$ (when A is the true model) and $\mathcal{O}_{BA}$ (when B is the true model), thus addressing how easy (or hard) it is to identify any of the models *if it is correct.* Thick lines mark the median odds, while the shaded areas encompass 90% of the realizations (i.e., they cover the range between the 5th and the 95th percentiles).

The odds ratio $\mathcal{O}$ increases roughly exponentially with the number of observations $N_{\rm obs}$, so our ability to distinguish between different models should rapidly improve in the coming years. Table II shows that in 5% of the realizations, as few as ~ 20 detections are enough to discriminate the 1g+1g population from the 2g+2g population at 5σ with Advanced LIGO at design sensitivity, while $N_{\rm obs} \sim 80$ observations are necessary to achieve 5σ confidence in 95% of the realizations.

Model selection involving the 1g+2g population typically requires a larger number of observations. This is clear when comparing the left panels of Figure 6 to the middle and right panels. In both the (1g+1g vs. 1g+2g) and (1g+2g vs. 2g+2g) comparisons the odds ratio grows (roughly) exponentially, but with smaller slope compared to the (1g+1g vs. 2g+2g) case. However the slope (and the odds ratio $\mathcal{O}$) is larger when 1g+2g is the true model: it is slightly easier to mistake a 1g+1g (or 2g+2g) population for a 1g+2g population than vice versa.

Model comparison is easier with more sensitive detectors. For example, distinguishing 1g+1g from 2g+2g at 5σ in 90% of the realizations requires only ~ 30 Voyager observations (instead of ~ 80 for Advanced LIGO at design sensitivity).

In Section V A, where only 3 observations were considered, the results were greatly dependent on the assumed mass distribution. Table II shows that this dependence becomes much weaker when more observations are available and/or the instrumental sensitivity improves. This is largely due to the discriminating power of the redshift distribution of the events, which becomes more relevant when high-z binaries become detectable (cf. Figure 5).

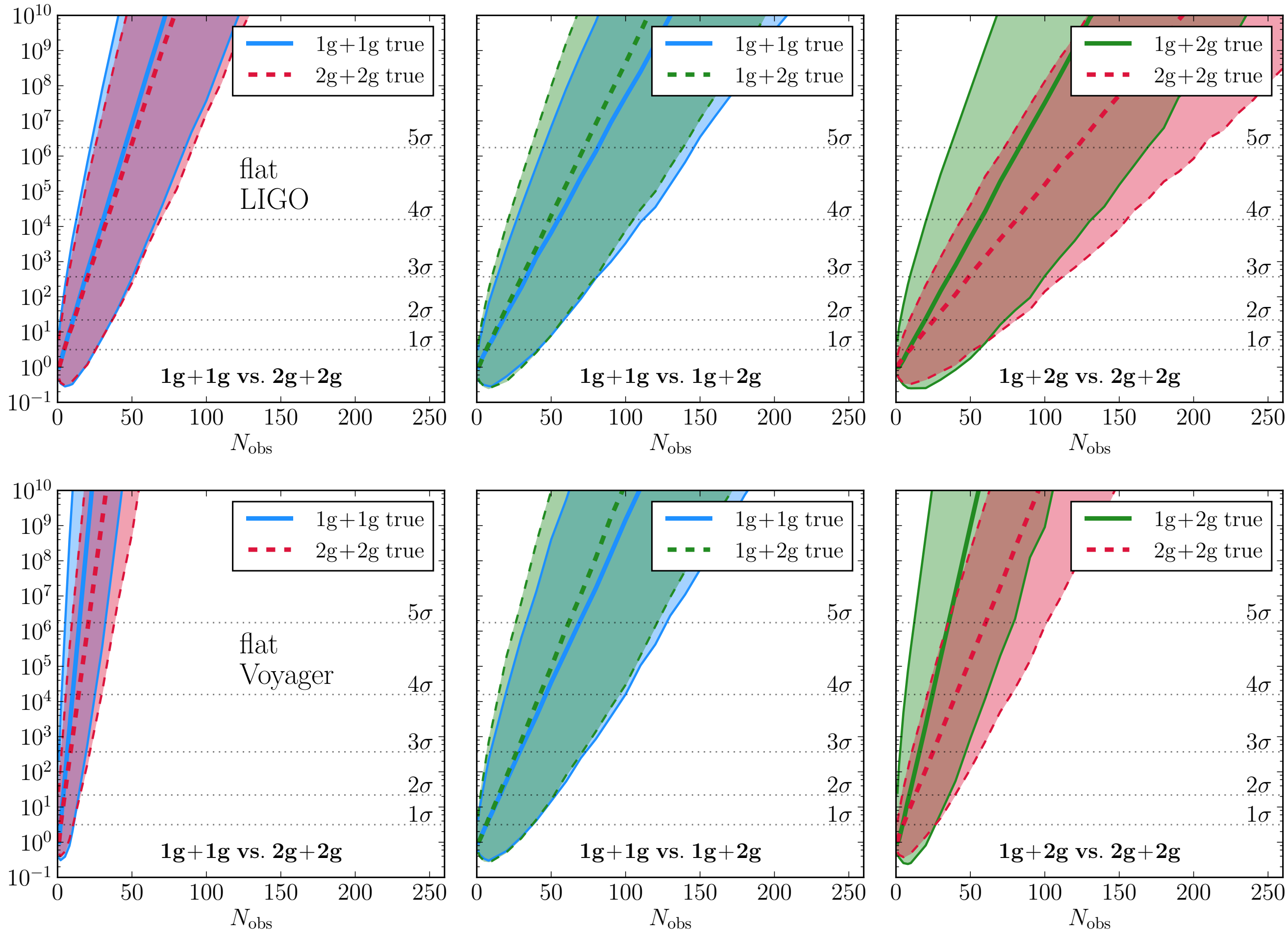

FIG. 6. Number of events that are necessary to distinguish populations for Advanced LIGO at design sensitivity (top) and Voyager (bottom). The median odds ratio (thick lines) and 90% confidence intervals to identify each model as true are plotted as functions of the number of observations $N_{\rm obs}$.

$N_{\rm obs}$ at 5σ		**1g1gT**/2g2g			**2g2gT**/1g1g			**1g1gT**/1g2g			**1g2gT**/1g1g			**1g2gT**/2g2g			**2g2gT**/1g2g		
		5%	50%	95%	5%	50%	95%	5%	50%	95%	5%	50%	95%	5%	50%	95%	5%	50%	95%
LIGO O1	flat	27	53	100	31	57	103	40	76	143	44	80	146	50	105	204	77	133	233
	log	27	52	94	25	50	94	30	58	106	29	56	106	42	86	165	59	104	182
	power law	14	29	57	19	35	64	7	17	34	13	23	41	31	61	114	35	64	117
Ad. LIGO	flat	23	46	86	26	50	91	45	82	146	37	73	139	37	83	170	73	122	206
	log	20	41	79	24	45	83	41	73	132	33	66	122	26	56	112	48	81	138
	power law	20	39	72	18	37	70	10	21	40	11	22	41	15	31	61	20	37	67
A+	flat	18	39	75	22	43	79	46	83	149	34	69	136	34	80	165	75	123	211
	log	16	34	65	19	38	69	41	73	131	30	62	120	22	51	107	50	81	136
	power law	17	35	67	17	34	65	10	22	41	10	21	40	12	27	52	20	35	61
Voyager	flat	6	15	33	10	21	39	34	69	128	27	62	122	13	36	80	36	61	102
	log	4	11	25	8	17	32	25	53	102	20	51	101	8	23	51	26	44	73
	power law	5	13	26	7	16	31	9	19	37	7	18	36	4	11	24	12	21	35

TABLE II. Number of observations needed to distinguish populations at 5σ with 5%, 50% and 95% probability. The "true" model is marked by a **T** in the column header. For instance, in column **1g1gT**/2g2g we compare models $1g + 1g$ and $2g + 2g$ when observations are drawn from the $1g + 1g$ catalog.

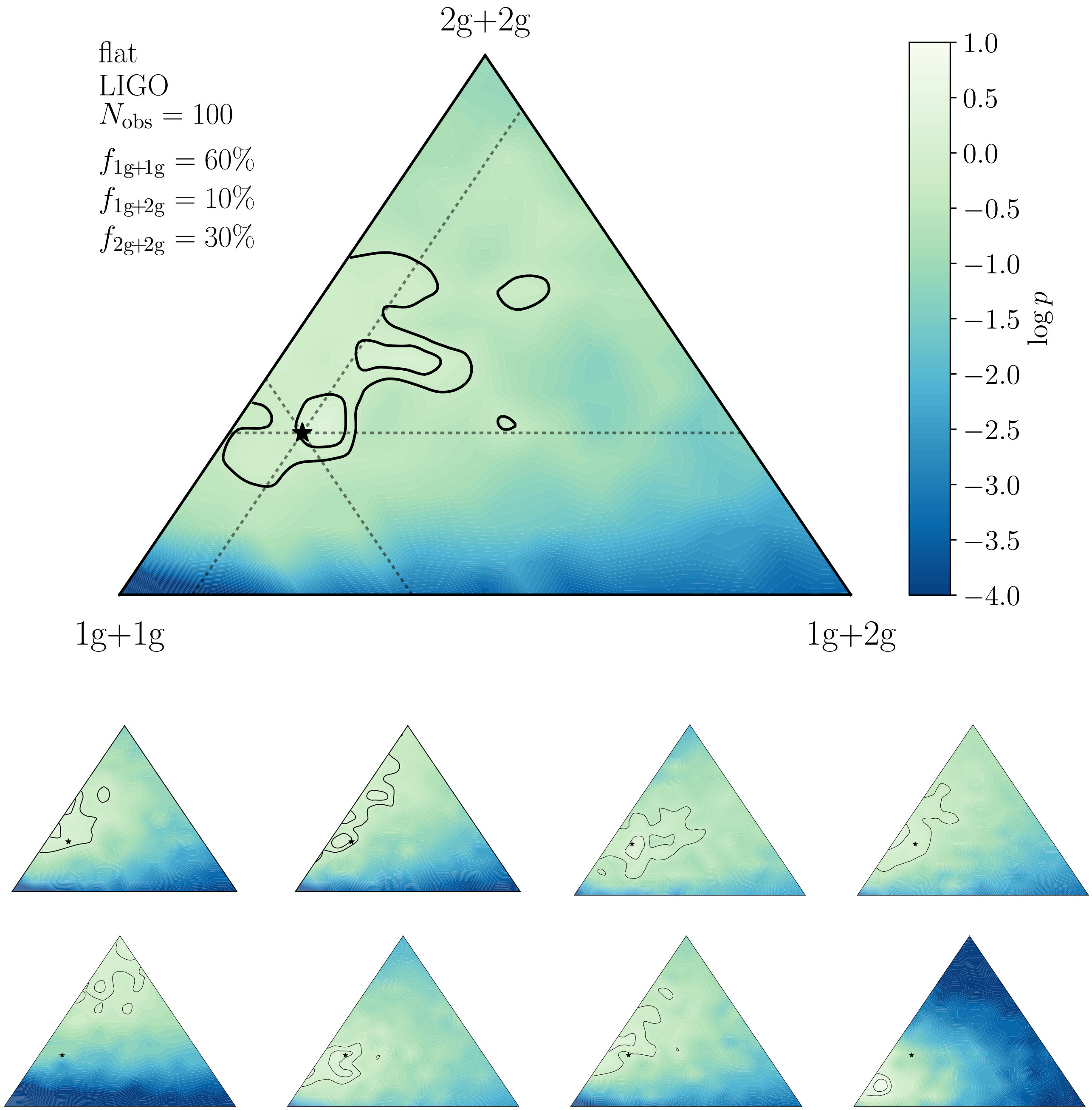

FIG. 7. Posterior distribution of the mixing fraction between the 1g+1g, 2g+2g and 1g+2g pure models. Each triangle shows the model space defined by $\sum f = 1$ for a given realization of $N_{\rm obs} = 100$ observations. The corners corresponds to the three pure models. The black star marks the "true" injected value of the mixing fractions. Each of the injected mixing fractions is constant along one of the dashed lines. The log-likelihood is shown in the color map: lighter regions are more likely than darker regions. Solid black contours mark the 50% and 90% confidence regions.

C. Simulated data: mixed models

Let us now turn to a more ambitious task. As anticipated in Section IV B, we now consider a population of binaries consisting of a *mixture* of the three pure models 1g+1g, 1g+2g and 2g+2g. The task is to measure their mixing fraction, i.e. to determine how many binaries belong to each of the three pure populations. This is computationally expensive, as it requires many evaluations of the likelihood defined in Eq. (26) through Monte Carlo methods.

As a proof of principle, in Figure 7 we show results for a specific choice of the mixing parameters. Simulated observations are drawn from a model[5] where 60% of the binaries are 1g+1g, 10% are 2g+2g, and 30% are 1g+2g:

$$\lambda_{\rm true} \equiv \left\{f_{\rm 1g+1g}, f_{\rm 1g+2g}, f_{\rm 2g+2g}\right\} = \left\{0.6, 0.1, 0.3\right\}. \quad (28)$$

For concreteness we assume the "flat" mass prescription and consider several realizations of $N_{\rm obs} = 100$ observations performed with the Advanced LIGO network at design sensitivity. Each of the triangles in Figure 7 shows the surface $f_{\rm 1g+1g}+f_{\rm 1g+2g}+f_{\rm 2g+2g} = 1$. The color coding corresponds to the values of the posterior $p(\lambda|\mathbf{d})$. Pure models lie on the corners of this "Dalitz plot", while the star marks the injected fraction.

As expected, measuring mixing fractions is sensibly harder than performing discrete model comparison, and it is going to require many more observations (a similar result was obtained in Ref. [90]). The injected fractions are recovered only in some of the realizations, suggesting that these data points are probably not enough to confidently perform the measurement.

In any case, we can note some trends. Most (but not all) of the realizations assign a rather low probability to the region where $f_{2g+2g} \sim 0$. Whenever a few 2g+2g events are present, their properties are sensibly different from those involving 1g BHs, and therefore the 2g+2g population can be identified relatively easily. Although we may be unlucky and estimate mixing fraction which are sensibly different from their true values, our model comparison algorithm does return a statistically consistent result. Out of 1000 realizations, we find that the correct mixing fraction is identified within the 50% (90%) confidence interval in 57% (25%) of the cases. The relatively small number of observations is likely to be one of the main reasons for this relatively pessimistic result: if we assume $N_{\rm obs} = 1000$, the correct mixing fraction is identified within the 50% (90%) confidence interval in 90% (77%) of the cases.

In conclusion, this preliminary study shows that measuring mixing fractions will be challenging in the near future. Estimating mixing fractions with high confidence may require several hundreds (if not thousands) of observations.

[5] Our injected fraction of 2g BHs was chosen only for illustrative purposes. It is higher than current estimates of merger rates in nuclear clusters, which are favorable environments for multiple merger events [32].

D. Caveats on mass functions and time delays

We have shown that, given a sufficient number of detected events, it is possible to distinguish a given 1g BH population from a variant of the *same* population where repeated mergers occur. Here we discuss how uncertainties in the assumed 1g mass distribution and in time delay prescriptions may bias our conclusions.

In Table III we perform pure model comparisons between BH binary populations that differ in *both* merger generation (1g+1g, 1g+2g, 2g+2g) scenario and in the assumed mass distribution. As shown in Section V B, the true distribution is correctly identified whenever it is among those tested. When the injected population is not among those being compared, differences in the assumed mass distribution can sometimes dominate over differences induced by the occurrence of subsequent mergers. For instance, if injected 2g+2g observations assuming the "flat" mass distributions are examined assuming a "power law" mass model, one would erroneously conclude that the observed population is 1g+1g, rather than 2g+2g.

However this should not be a problem in practice, because the mass distribution should soon be well constrained by observations. Realistic astrophysical scenarios typically predict a small fraction of multiple mergers, i.e., a small fraction of 2g events. Even remaining theory-agnostic, this anticipation is already (although inconclusively) supported by the data. Our Table I suggests that 1g+1g mergers may already be favored over 2g scenarios. So, in practice, there are theoretical and (hopefully soon) experimental reason to assume that the majority of detected events have a 1g+1g origin. In this very plausible scenario, the model selection procedure can be "bootstrapped" as follows:

(i) The mass distribution is inferred from a large enough number of detections, assuming that most events are 1g+1g;

(ii) This observationally inferred mass distribution can be used to replace our "toy" mass distributions (flat, log or power law) for 1g+1g BHs, and the 2g distributions can be constructed through hierarchical mergers as described earlier;

(iii) We can now look at all measurable properties of the population to determine whether some (presumably small) fraction of events has a 2g origin.

Table III shows that step (i) above does not present problems. Indeed, according to Table III, while it is indeed possible to wrongly rule in favor of 1g+1g BHs given

Injection	Test			Preferred	
flat 1g+1g	flat 1g+1g	vs.	flat 2g+2g	flat 1g+1g	✓T
	log 1g+1g	vs.	log 2g+2g	not significant	
	power law 1g+1g	vs.	power law 2g+2g	power law 1g+1g	✓
flat 2g+2g	flat 1g+1g	vs.	flat 2g+2g	flat 2g+2g	✓T
	log 1g+1g	vs.	log 2g+2g	log 2g+2g	✓
	power law 1g+1g	vs.	power law 2g+2g	power law 1g+1g	✗
log 1g+1g	flat 1g+1g	vs.	flat 2g+2g	flat 1g+1g	✓
	log 1g+1g	vs.	log 2g+2g	log 1g+1g	✓T
	power law 1g+1g	vs.	power law 2g+2g	power law 1g+1g	✓
log 2g+2g	flat 1g+1g	vs.	flat 2g+2g	flat 1g+1g	✗
	log 1g+1g	vs.	log 2g+2g	log 2g+2g	✓
	power law 1g+1g	vs.	power law 2g+2g	power law 1g+1g	✗
power law 1g+1g	flat 1g+1g	vs.	flat 2g+2g	flat 1g+1g	✓
	log 1g+1g	vs.	log 2g+2g	log 1g+1g	✓
	power law 1g+1g	vs.	power law 2g+2g	power law 1g+1g	✓T
power law 2g+2g	flat 1g+1g	vs.	flat 2g+2g	flat 1g+1g	✗
	log 1g+1g	vs.	log 2g+2g	log 1g+1g	✗
	power law 1g+1g	vs.	power law 2g+2g	power law 2g+2g	✓T

TABLE III. Model comparison tests between populations characterized by different merger generations *and* mass distributions using the Advanced LIGO sensitivity curve. For each injected distribution and model comparison we report the preferred population in the limit where $N_{\rm obs} \to \infty$ (in practice we use $N_{\rm obs} = 10^3$). The true population (T) is correctly identified whenever it is among those tested. While most of the comparisons correctly identify the merger generation (rows denoted by a ✓check mark), in some cases making the wrong assumption on the underlying mass distribution prevents a correct identification (rows denoted by a ✗). In one case (second row) we obtained odds ratios consistent with one even when $N_{\rm obs} \to \infty$, so that no conclusions can be drawn and the comparison is marked as "not significant". In all other cases the behavior of $\mathcal{O}(N_{\rm obs})$ is qualitatively similar to Fig. 6, i.e. the odds ratio grows exponentially until populations can be distinguished at 5σ.

2g+2g injections, the converse is unlikely: if model selection favors 2g+2g BHs, the injected data never belong to a 1g+1g population with a different mass spectrum.

To quantify the importance of time delays, we repeated all the comparisons shown in Table II excluding redshift information, i.e. taking $\mathbf{u} = \{M, q, \chi_{\rm eff}\}$ as our vector of observable quantities. We find that the correct population is always identified. The odds still grow exponentially with the number of observations, although with somewhat shallower slopes. This is expected, because the statistical analysis is performed using less information. Omitting redshift information does not significantly affect the 1g+1g vs. 2g+2g and 1g+2g vs. 2g+2g comparisons, but it plays a more important role in the 1g+1g vs 1g+2g comparisons. This is because (as illustrated in Fig. 5) the mass distributions are very similar for these populations, which are therefore harder to distinguish if redshifts are ignored. For instance, while Fig. 6 shows that $\sim$ 40 observations are enough to distinguish the 1g+1g and 1g+2g "flat" populations at 3σ with LIGO in 50% of the realizations, up to $\sim$ 200 events will be necessary to reach the same conclusion in the absence of redshift information.

VI. CONCLUSIONS

The main result of this paper is that GW observations of merging stellar-mass BH binaries can be used to gather information about their progenitors. Starting from simple, physically motivated populations of "first generation" (1g) BHs born from stellar collapse, we construct populations where merging binaries include "second generation" (2g) BHs, whose masses and spins are computed using numerical relativity fitting formulas. Then we use Bayesian model selection to determine whether current or future ground-based GW interferometers can distinguish different populations. If 2g BHs occur in nature, it should be possible to recover evidence for their existence from GW data; otherwise, the data can be used to constrain astrophysical models that produce 2g BHs.

As a first application of our Bayesian model selection framework, we perform model selection using the two confirmed detections (GW150914 and GW151226) and the LVT151012 trigger from Advanced LIGO's first observing run. It is quite remarkable that, even with only three data points, some of the comparisons show odds ratios as high as $\sim$ 10 in favor of 1g BHs. As expected, model selection performance improves with more observations and more sensitive detectors. Indeed, as shown in Figure 6, the Bayesian odds ratio for comparisons between two pure models scales (roughly) exponentially with the number of observations. Depending on the actual realization, $\sim$ 20–200 Advanced LIGO observations at design sensitivity should allow us to discriminate which of the three populations is favored by the data at 5σ confidence level in one-to-one comparisons. Instrumental upgrades will bring this number down to 15–200 observations for A+, and 5–100 for Voyager.

More realistically, astrophysical populations of merging binaries will be a mixture of all three populations (1g+1g, 1g+2g, 2g+2g), and the real experimental task is to determine the relative mixing fractions. Using simulated data, we construct synthetic catalogs assuming a mixture of models for the different BH generations, and attempt to measure the mixing fractions using Bayesian inference. Our preliminary results suggest that this is a much more challenging task: recovering the mixing fractions may require several hundreds (if not thousands) of observations.

This work should be regarded as a proof-of-principle study that can (and should) be extended in several directions. Our simple models are not supposed to be astrophysically realistic: they were developed solely to show that, at least in principle, GW observations could provide information on the occurrence of multiple stellar-mass BH mergers. The inclusion of detailed spin alignment models and more realistic mass distributions (see e.g. [94]), preferably with input from population synthesis codes, is an important topic for future investigation.

As illustrated in Section II E (see in particular Figure 3), the spin magnitudes of the merging BHs are very sensitive to their merger history. This is also true for the massive BH binaries observable by LISA: see e.g. [40, 95]. Unlike BHs born from stellar collapse, the spin distribution of post-merger BHs should be strongly peaked at $\chi_{\rm f} \sim 0.7$. In this paper we only considered measurements of the "effective spin" $\chi_{\rm eff}$, because this is the spin parameter that enters at lowest PN order in the gravitational waveform. This is a very conservative approach. As shown in Figure 2, the "memory effect" encoded in the spin magnitudes is largely washed out in this variable. Measurements of the individual spin magnitudes should be possible by considering better waveform models or higher SNR signals: in this sense, our predictions should be regarded as conservative. Moreover, high-SNR ringdown observations will allow measurements of the final (post-merger) spin $\chi_{\rm f}$ within a few percent [69]. These measurements could also be used to identify the progenitors of merging BHs.

The model selection framework developed in this paper is complementary to other studies, which usually focus on discriminating specific astrophysical formation channels (e.g., field binaries vs. dynamical formation scenarios [48, 91, 96–99], but see also [100] for work on intermediate-mass BHs). We focused on using statistical distributions consisting of several observations, but it is possible that single events may be smoking guns for (or against) multiple merger scenarios, at the price of making stronger assumptions on the formation mechanism of 1g BHs. For example, binaries with component masses above the pair-instability gap [44] can point to the occurrence of multiple mergers is we assume that 1g BHs are formed via core collapse, and if we are confident about the upper mass limit on 1g BHs set by pair instabilities. We hope that our approach will spark more studies of the astrophysical information encoded in present and future GW data sets.

While completing our study we learned that Maya Fishbach, Daniel Holz and Ben Farr have been pursuing a similar investigation [101]. Their work nicely complements our study, as they focus on the spin distributions and address the detectability of more than two generations of mergers.

ACKNOWLEDGMENTS

We thank Archisman Ghosh, Walter Del Pozzo and Parameswaran Ajith for sharing parameter estimation data from Ref. [77]. We also thank Daniel Holz, Maya Fishbach, Ben Farr, Leo Stein and Chris Moore for discussions. DG is supported by NASA through Einstein Postdoctoral Fellowship grant No. PF6-170152 awarded by the Chandra X-ray Center, which is operated by the Smithsonian Astrophysical Observatory for NASA under contract NAS8-03060. EB was supported by NSF Grant No. PHY-1607130 and by FCT contract IF/00797/2014/CP1214/CT0012 under the IF2014 Programme. This work was supported by the H2020-MSCA-RISE-2015 Grant No. StronGrHEP-690904. Computations were performed on the Caltech computer cluster "Wheeler," supported by the Sherman Fairchild Foundation and Caltech. Partial support is acknowledged by NSF Award CAREER PHY-1151197. This research made use of python packages ASTROPY [102] and MATPLOTLIB [103].

[1] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), PRL **116**, 061102 (2016), arXiv:1602.03837 [gr-qc].

[2] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), PRL **116**, 241103 (2016), arXiv:1606.04855 [gr-qc].

[3] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Physical Review X **6**, 041015 (2016), arXiv:1606.04856 [gr-qc].

[4] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), ApJ **818**, L22 (2016), arXiv:1602.03846 [astro-ph.HE].

[5] K. A. Postnov and L. R. Yungelson, Living Reviews in Relativity **17**, 3 (2014), arXiv:1403.4754 [astro-ph.HE].

[6] M. J. Benacquista and J. M. B. Downing, Living Reviews in Relativity **16**, 4 (2013), arXiv:1110.4423 [astro-ph.SR].

[7] M. Mapelli, M. Colpi, and L. Zampieri, MNRAS **395**, L71 (2009), arXiv:0902.3540 [astro-ph.HE].

[8] K. Belczynski, M. Dominik, T. Bulik, R. O'Shaughnessy, C. Fryer, and D. E. Holz, ApJ **715**, L138 (2010), arXiv:1004.0386 [astro-ph.HE].

[9] M. Dominik, K. Belczynski, C. Fryer, D. E. Holz, E. Berti, T. Bulik, I. Mandel, and R. O'Shaughnessy, ApJ **759**, 52 (2012), arXiv:1202.4901 [astro-ph.HE].

[10] M. Dominik, K. Belczynski, C. Fryer, D. E. Holz, E. Berti, T. Bulik, I. Mandel, and R. O'Shaughnessy, ApJ **779**, 72 (2013), arXiv:1308.1546 [astro-ph.HE].

[11] M. Spera, M. Mapelli, and A. Bressan, MNRAS **451**, 4086 (2015), arXiv:1505.05201 [astro-ph.SR].

[12] M. Dominik, E. Berti, R. O'Shaughnessy, I. Mandel, K. Belczynski, C. Fryer, D. E. Holz, T. Bulik, and F. Pannarale, ApJ **806**, 263 (2015), arXiv:1405.7016 [astro-ph.HE].

[13] K. Belczynski, D. E. Holz, T. Bulik, and R. O'Shaughnessy, Nature **534**, 512 (2016), arXiv:1602.04531 [astro-ph.HE].

[14] K. Belczynski, S. Repetto, D. E. Holz, R. O'Shaughnessy, T. Bulik, E. Berti, C. Fryer, and M. Dominik, ApJ **819**, 108 (2016), arXiv:1510.04615 [astro-ph.HE].

[15] C. L. Rodriguez, M. Morscher, B. Pattabiraman, S. Chatterjee, C.-J. Haster, and F. A. Rasio, PRL **115**, 051101 (2015), arXiv:1505.00792 [astro-ph.HE].

[16] C. L. Rodriguez, S. Chatterjee, and F. A. Rasio, PRD **93**, 084029 (2016), arXiv:1602.02444 [astro-ph.HE].

[17] S. Chatterjee, C. L. Rodriguez, and F. A. Rasio, ApJ **834**, 68 (2017), arXiv:1603.00884.

[18] C. L. Rodriguez, C.-J. Haster, S. Chatterjee, V. Kalogera, and F. A. Rasio, ApJ **824**, L8 (2016), arXiv:1604.04254 [astro-ph.HE].

[19] R. A. Remillard and J. E. McClintock, ARA&A **44**, 49 (2006), astro-ph/0606352.

[20] R. Narayan and J. E. McClintock, ArXiv e-prints (2013), arXiv:1312.6698 [astro-ph.HE].

[21] L. Wen, ApJ **598**, 419 (2003), astro-ph/0211492.

[22] F. Antonini, N. Murray, and S. Mikkola, ApJ **781**, 45 (2014), arXiv:1308.3674 [astro-ph.HE].

[23] F. Antonini, S. Chatterjee, C. L. Rodriguez, M. Morscher, B. Pattabiraman, V. Kalogera, and F. A. Rasio, ApJ **816**, 65 (2016), arXiv:1509.05080.

[24] T. Kinugawa, K. Inayoshi, K. Hotokezaka, D. Nakauchi, and T. Nakamura, MNRAS **442**, 2963 (2014), arXiv:1402.6672 [astro-ph.HE].

[25] T. Hartwig, M. Volonteri, V. Bromm, R. S. Klessen, E. Barausse, M. Magg, and A. Stacy, MNRAS **460**, L74 (2016), arXiv:1603.05655.

[26] K. Belczynski, T. Ryu, R. Perna, E. Berti, T. L. Tanaka, and T. Bulik, ArXiv e-prints (2016), arXiv:1612.01524 [astro-ph.HE].

[27] I. Mandel and S. E. de Mink, MNRAS **458**, 2634 (2016), arXiv:1601.00007 [astro-ph.HE].

[28] S. E. de Mink and I. Mandel, MNRAS **460**, 3545 (2016), arXiv:1603.02291 [astro-ph.HE].

[29] P. Marchant, N. Langer, P. Podsiadlowski, T. M. Tauris, and T. J. Moriya, A&A **588**, A50 (2016), arXiv:1601.03718 [astro-ph.SR].

[30] S. Bird, I. Cholis, J. B. Muñoz, Y. Ali-Haïmoud, M. Kamionkowski, E. D. Kovetz, A. Raccanelli, and A. G. Riess, PRL **116**, 201301 (2016), arXiv:1603.00464.

[31] I. Cholis, E. D. Kovetz, Y. Ali-Haïmoud, S. Bird, M. Kamionkowski, J. B. Muñoz, and A. Raccanelli, PRD **94**, 084013 (2016), arXiv:1606.07437 [astro-ph.HE].

[32] F. Antonini and F. A. Rasio, ApJ **831**, 187 (2016), arXiv:1606.04889 [astro-ph.HE].

[33] M. Mapelli, MNRAS **459**, 3432 (2016), arXiv:1604.03559.

[34] R. M. O'Leary, Y. Meiron, and B. Kocsis, ApJ **824**, L12 (2016), arXiv:1602.02809 [astro-ph.HE].

[35] D. Merritt, M. Milosavljević, M. Favata, S. A. Hughes, and D. E. Holz, ApJ **607**, L9 (2004), astro-ph/0402057.

[36] 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, ArXiv e-prints (2017), arXiv:1702.07818 [astro-ph.HE].

[37] J. M. Bellovary, M.-M. Mac Low, B. McKernan, and K. E. S. Ford, ApJ **819**, L17 (2016), arXiv:1511.00005.

[38] S. Clesse and J. García-Bellido, Physics of the Dark Universe **15**, 142 (2017), arXiv:1603.05234.

[39] D. Kushnir, M. Zaldarriaga, J. A. Kollmeier, and R. Waldman, MNRAS **462**, 844 (2016), arXiv:1605.03839 [astro-ph.HE].

[40] E. Berti and M. Volonteri, ApJ **684**, 822-828 (2008), arXiv:0802.0025.

[41] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), ArXiv e-prints (2017), arXiv:1704.04628 [gr-qc].

[42] A. Heger and S. E. Woosley, ApJ **567**, 532 (2002), astro-ph/0107037.

[43] S. E. Woosley, ApJ **836**, 244 (2017), arXiv:1608.08939 [astro-ph.HE].

[44] 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, A&A **594**, A97 (2016), arXiv:1607.03116 [astro-ph.HE].

[45] K. Belczynski, R. E. Taam, E. Rantsiou, and M. van der Sluys, ApJ **682**, 474-486 (2008), astro-ph/0703131.

[46] M. C. Miller and J. M. Miller, Phys. Rep. **548**, 1 (2015), arXiv:1408.4145 [astro-ph.HE].

[47] K. Belczynski, D. E. Holz, T. Bulik, and R. O'Shaughnessy, Nature **534**, 512 (2016), arXiv:1602.04531 [astro-ph.HE].

[48] C. L. Rodriguez, M. Zevin, C. Pankow, V. Kalogera, and F. A. Rasio, ApJ **832**, L2 (2016), arXiv:1609.05916 [astro-ph.HE].

[49] V. Kalogera, ApJ **541**, 319 (2000), astro-ph/9911417.

[50] D. Gerosa, M. Kesden, E. Berti, R. O'Shaughnessy, and U. Sperhake, PRD **87**, 104028 (2013), arXiv:1302.4442 [gr-qc].

[51] T. Bogdanović, C. S. Reynolds, and M. C. Miller, ApJ **661**, L147 (2007), astro-ph/0703054.

[52] D. Gerosa, M. Kesden, U. Sperhake, E. Berti, and R. O'Shaughnessy, PRD **92**, 064016 (2015), arXiv:1506.03492 [gr-qc].

[53] M. Kesden, D. Gerosa, R. O'Shaughnessy, E. Berti, and U. Sperhake, PRL **114**, 081103 (2015), arXiv:1411.0674 [gr-qc].

[54] D. Gerosa and M. Kesden, PRD **93**, 124066 (2016), arXiv:1605.01067 [astro-ph.HE].
[55] E. Barausse, V. Morozova, and L. Rezzolla, ApJ **758**, 63 (2012), arXiv:1206.3803 [gr-qc].
[56] E. Barausse and L. Rezzolla, ApJ **704**, L40 (2009), arXiv:0904.2577 [gr-qc].
[57] C. O. Lousto and Y. Zlochower, PRD **89**, 104052 (2014), arXiv:1312.5775 [gr-qc].
[58] J. Healy, C. O. Lousto, and Y. Zlochower, PRD **90**, 104004 (2014), arXiv:1406.7295 [gr-qc].
[59] Y. Zlochower and C. O. Lousto, PRD **92**, 024022 (2015), arXiv:1503.07536 [gr-qc].
[60] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, PRD **93**, 044006 (2016), arXiv:1508.07250 [gr-qc].
[61] F. Hofmann, E. Barausse, and L. Rezzolla, ApJ **825**, L19 (2016), arXiv:1605.01938 [gr-qc].
[62] J. Healy and C. O. Lousto, PRD **95**, 024037 (2017), arXiv:1610.09713 [gr-qc].
[63] X. Jiménez-Forteza, D. Keitel, S. Husa, M. Hannam, S. Khan, and M. Pürrer, PRD **95**, 064024 (2017), arXiv:1611.00332 [gr-qc].
[64] D. W. Hogg, ArXiv Astrophysics e-prints (1999), astro-ph/9905116.
[65] P. A. R. Ade *et al.* (Planck Collaboration), A&A **594**, A13 (2016), arXiv:1502.01589.
[66] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), PRL **116**, 241102 (2016), arXiv:1602.03840 [gr-qc].
[67] É. Racine, PRD **78**, 044021 (2008), arXiv:0803.1820 [gr-qc].
[68] M. Pürrer, M. Hannam, and F. Ohme, PRD **93**, 084042 (2016), arXiv:1512.04955 [gr-qc].
[69] E. Berti, A. Sesana, E. Barausse, V. Cardoso, and K. Belczynski, PRL **117**, 101102 (2016), arXiv:1605.09286 [gr-qc].
[70] S. Vitale, PRD **94**, 121501 (2016), arXiv:1610.06914 [gr-qc].
[71] S. Vitale and M. Evans, ArXiv e-prints (2016), arXiv:1610.06917 [gr-qc].
[72] S. Vitale, R. Lynch, V. Raymond, R. Sturani, J. Veitch, and P. Graff, PRD **95**, 064053 (2017), arXiv:1611.01122 [gr-qc].
[73] L. Santamaría, F. Ohme, P. Ajith, B. Brügmann, N. Dorband, M. Hannam, S. Husa, P. Mösta, D. Pollney, C. Reisswig, E. L. Robinson, J. Seiler, and B. Krishnan, PRD **82**, 064016 (2010), arXiv:1005.3306 [gr-qc].
[74] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Living Reviews in Relativity **19**, 1 (2016), arXiv:1304.0670 [gr-qc].
[75] D. Mc Clelland, M. Evans, R. Schnabel, B. Lantz, I. Martin, and V. Quetschke, LIGO Document Control Center `https://dcc.ligo.org/LIGO-T1400316/public` (2015).
[76] J. Abadie *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), CQG **27**, 173001 (2010), arXiv:1003.2480 [astro-ph.HE].
[77] A. Ghosh, W. Del Pozzo, and P. Ajith, PRD **94**, 104070 (2016), arXiv:1505.05607 [gr-qc].
[78] 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, PRD **91**, 042003 (2015), arXiv:1409.7215 [gr-qc].
[79] P. Ajith and S. Bose, PRD **79**, 084032 (2009), arXiv:0901.4936 [gr-qc].
[80] T. B. Littenberg, J. G. Baker, A. Buonanno, and B. J. Kelly, PRD **87**, 104003 (2013), arXiv:1210.0893 [gr-qc].
[81] H.-S. Cho, E. Ochsner, R. O'Shaughnessy, C. Kim, and C.-H. Lee, PRD **87**, 024004 (2013), arXiv:1209.4494 [gr-qc].
[82] P. B. Graff, A. Buonanno, and B. S. Sathyaprakash, PRD **92**, 022002 (2015), arXiv:1504.04766 [gr-qc].
[83] E. Poisson and C. M. Will, PRD **52**, 848 (1995), gr-qc/9502040.
[84] S. A. Hughes, MNRAS **331**, 805 (2002), astro-ph/0108483.
[85] E. Berti, A. Buonanno, and C. M. Will, PRD **71**, 084025 (2005), gr-qc/0411129.
[86] E. Di Valentino, A. Melchiorri, and J. Silk, Physics Letters B **761**, 242 (2016), arXiv:1606.00634.
[87] A. G. Riess, L. M. Macri, S. L. Hoffmann, D. Scolnic, S. Casertano, A. V. Filippenko, B. E. Tucker, M. J. Reid, D. O. Jones, J. M. Silverman, R. Chornock, P. Challis, W. Yuan, P. J. Brown, and R. J. Foley, ApJ **826**, 56 (2016), arXiv:1604.01424.
[88] J. L. Bernal, L. Verde, and A. G. Riess, J. Cosmology Astropart. Phys. **10**, 019 (2016), arXiv:1607.05617.
[89] J. R. Gair, C. Tang, and M. Volonteri, PRD **81**, 104014 (2010), arXiv:1004.1921 [astro-ph.GA].
[90] A. Sesana, J. Gair, E. Berti, and M. Volonteri, PRD **83**, 044036 (2011), arXiv:1011.5893 [astro-ph.CO].
[91] S. Stevenson, F. Ohme, and S. Fairhurst, ApJ **810**, 58 (2015), arXiv:1504.07802 [astro-ph.HE].
[92] J. R. Gair, A. Sesana, E. Berti, and M. Volonteri, CQG **28**, 094018 (2011), arXiv:1009.6172 [gr-qc].
[93] E. Berti, J. Gair, and A. Sesana, PRD **84**, 101501 (2011), arXiv:1107.3528 [gr-qc].
[94] E. D. Kovetz, I. Cholis, P. C. Breysse, and M. Kamionkowski, ArXiv e-prints (2016), arXiv:1611.01157.
[95] A. Sesana, E. Barausse, M. Dotti, and E. M. Rossi, ApJ **794**, 104 (2014), arXiv:1402.7088.
[96] S. Vitale, R. Lynch, R. Sturani, and P. Graff, CQG **34**, 03LT01 (2017), arXiv:1503.04307 [gr-qc].
[97] A. Nishizawa, E. Berti, A. Klein, and A. Sesana, PRD **94**, 064020 (2016), arXiv:1605.01341 [gr-qc].
[98] K. Breivik, C. L. Rodriguez, S. L. Larson, V. Kalogera, and F. A. Rasio, ApJ **830**, L18 (2016), arXiv:1606.09558.
[99] A. Nishizawa, A. Sesana, E. Berti, and A. Klein, MNRAS **465**, 4375 (2017), arXiv:1606.09295 [astro-ph.HE].
[100] C.-J. Haster, Z. Wang, C. P. L. Berry, S. Stevenson, J. Veitch, and I. Mandel, MNRAS **457**, 4499 (2016), arXiv:1511.01431 [astro-ph.HE].
[101] M. Fishbach, D. E. Holz, and B. Farr, Astrophys. J. **840**, L24 (2017), arXiv:1703.06869 [astro-ph.HE].
[102] T. P. Robitaille *et al.* (Astropy Collaboration), A&A **558**, A33 (2013), arXiv:1307.6212 [astro-ph.IM].
[103] J. D. Hunter, Computing in Science and Engineering **9**, 90 (2007).