Bayesian analysis of LIGO-Virgo mergers: Primordial versus astrophysical black hole populations

Alex Hall,[1,*] Andrew D. Gow,[2] and Christian T. Byrnes[2]

[1]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, United Kingdom*

[2]*Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, United Kingdom*

(Received 8 September 2020; accepted 12 November 2020; published 11 December 2020)

We conduct a thorough Bayesian analysis of the possibility that the black hole merger events seen in gravitational waves are primordial black hole (PBH) mergers. Using the latest merger rate models for PBH binaries drawn from a log-normal mass function, we compute posterior parameter constraints and Bayesian evidences using data from the first two observing runs of LIGO-Virgo. We account for theoretical uncertainty due to possible disruption of the binary by surrounding PBHs, which can suppress the merger rate significantly. We also consider simple astrophysically motivated models and find that these are favored decisively over the PBH scenario, quantified by the Bayesian evidence ratio. Paying careful attention to the influence of the parameter priors and the quality of the model fits, we show that the evidence ratios can be understood by comparing the predicted chirp mass distribution to that of the data. We identify the posterior predictive distribution of chirp mass as a vital tool for discriminating between models. A model in which all mergers are PBH binaries is strongly disfavored compared with astrophysical models, in part due to the overprediction of heavy systems having $\mathcal{M}_{\rm chirp} \gtrsim 40\ M_\odot$ and positive skewness over the range of observed masses which does not match the observations. We find that the fit is not significantly improved by adding a maximum mass cutoff or a bimodal mass function or imposing that PBH binaries form at late times. We argue that a successful PBH model must either modify the log-normal shape of the initial mass function significantly or abandon the hypothesis that all observed merging binaries are primordial. We develop and apply techniques for analyzing PBH models with gravitational wave data, which will be necessary for robust statistical inference as the gravitational wave source sample size increases.

DOI: 10.1103/PhysRevD.102.123524

I. INTRODUCTION

Primordial black holes (PBHs) [1–4] have long been recognized as a unique dark matter candidate that does not require the existence of a new particle or modification to gravity (see Refs. [5–8] for recent reviews). Interest in PBHs has increased greatly due to the detection of black hole (BH) mergers emitting gravitational waves (GWs) by LIGO and Virgo [9], since it is possible that the merging objects are primordial in origin [10–12].

Assuming that some fraction of the observed merger events are primordial binaries, one can place bounds on the fraction of the dark matter that should be in PBHs of the relevant mass range to explain the observed merger rate (see Ref. [13] for a review). If all of the confirmed LIGO-Virgo events are PBH mergers, then the fraction of dark matter in PBHs, $f_{\rm PBH}$, is typically found to be of order 10^{-3} depending on assumptions about the evolution and formation mechanism of the binary (see Ref. [14] for a comprehensive recent review), although $f_{\rm PBH} \approx 1$ is still permitted in certain models (e.g., Ref. [15]).

The increase in sample size to ten events [16] since the first detection has allowed several groups to make fits of the PBH initial mass function to the LIGO data, typically concluding that a log-normal mass function with central mass $m_c \sim 10\ M_\odot$ and a width of order unity is the best fit [17–19]. Connecting the empirical distribution of black hole source parameters to an initial mass function for PBHs is in general nontrivial and involves modeling the formation of the binary and its evolution through to the merger event [20,21], but the reward for this is a direct constraint on the conditions in the early Universe which gave rise to PBH formation. The initial mass function can be predicted from the spectrum of curvature fluctuations at the formation epoch, implying that constraints on the mass function can give unique information on a host of poorly understood physics in the early Universe, including the small-scale power spectrum, non-Gaussianity, phase transitions, and inflation [22–25]. In general, the calculation is more complicated for the extended mass functions required by the LIGO data if more than one event is primordial, but,

*ahall@roe.ac.uk

encouragingly, recent simplified models for PBH binary evolution tentatively give good agreement with the results of N-body simulations [26].

What previous analyses have neglected, however, is whether the best-fit PBH model is a good fit to the data and, more specifically, whether the model is a good fit compared to simple astrophysical BH merger models such as those studied by the LIGO and Virgo collaborations. With the event rate in the recent LIGO-Virgo O3 observing run roughly double what it was in the O1 and O2 observing runs, the importance of a rigorous statistical analysis of the PBH formation scenario is becoming increasingly necessary. The required analysis can be compared with more conventional studies of stellar black hole binary populations using GW events [27–31] where techniques such as Bayesian model comparison, tests of model consistency, and goodness-of-fit tests are becoming commonplace. These tests are in principle able to rule out whole classes of PBH mass functions for *any* values of their parameters, if those mass functions predict merger populations which do not match the observations. The ability of Bayesian methods to quantify this is one of several advantages to pursuing this line of study.

In this paper, we perform Bayesian tests of the PBH merger scenario using the binary black hole (BBH) merger events in the first two observing runs of LIGO-Virgo. We quantify how well the data fit the PBH merger scenario compared with simple astrophysically motivated models using the latest calculations for the formation of binaries during radiation domination and their subsequent evolution and possible disruption. The techniques we employ provide a link between the methodology of the LIGO-Virgo and GW community and that of the PBH community. We consider individual source masses and redshifts in our analysis, which provides more constraining power than simply using the component spins (for which a Bayesian analysis in the spirit of ours was performed recently in Ref. [32]). By accounting for correlated parameter uncertainties, the nonuniform selection probability of LIGO-Virgo, and an accurate likelihood function for the population parameters, we provide a comprehensive statistical study of the PBH formation channel for merging black hole binaries. We pay particular attention to how the models are able to fit the data and why certain models are favored over others.

As is well known, Bayesian tests using the model posterior probability or evidence are sensitive to the priors assigned to the parameters of each model, which often lack a strong physical motivation when the models are phenomenological. In this case, the best one can do is transparently present the chosen priors and check the sensitivity of the results to alternative choices. The sensitivity is typically only logarithmic, but we will be careful to account for uncertainty in the choice of prior when presenting our results.

Unless otherwise stated, we adopt units where $G = c = 1$. When computing background quantities, we assume a flat ΛCDM cosmological model with parameters fixed to the best-fitting values of Planck 2015 [33]. We note that the sources considered in this work are all at sufficiently low redshift that our results are insensitive to the choice of cosmological model.

II. DATA

To test PBH models of binary mergers, we use the ten BH-BH merger events in the Gravitational-Wave Transient Catalogue from the first two observing runs of LIGO-Virgo (GWTC-1) [16].[1] As we shall see, it is sufficient for our population-level analysis to use only samples from the posterior distribution of source parameters (masses, spins, etc.) for each source. In Fig. 1, we show posterior samples in the plane of detector-frame (i.e., redshifted) chirp mass $\mathcal{M}_z$ and mass ratio q, where $\mathcal{M}_z = (1+z)(m_1 m_2)^{3/5}(m_1+m_2)^{-1/5}$ and $q = m_2/m_1$. Note that $m_2 < m_1$ has been enforced in the GWTC-1 posteriors.

The detector-frame chirp mass is a well-constrained parameter for each source, being constrained with typically less than or approximately equal to 10% precision. Being closely related to measurable aspects of the source waveforms (specifically the frequency evolution of the GW strain), it is also practically uncorrelated with other source parameters, as shown for the case of q in Fig. 1. This is in contrast with other descriptors of the absolute mass scale of the BBH; the heavier mass m_1 is typically only constrained at the 10%–50% level and is highly correlated with the lighter mass m_2, while the total mass is on average constrained with greater than or approximately equal to 10% precision and is often highly correlated with q. For this reason, we will often present constraints in terms of chirp mass rather than heaviest mass or total mass.[2] Note that the redshifts of the sources in GWTC-1 are sufficiently low that there is only a small difference between source-frame and detector-frame masses.

Any inference of BBH populations must carefully account for the selection function of LIGO-Virgo. We adopt an accurate semianalytic approach to computing $p_{\rm det}(\boldsymbol{\lambda})$, the probability of detecting a source given it has source parameters $\boldsymbol{\lambda}$, following the prescription in Refs. [27,31]. We use the public code GWDET [35] to compute $p_{\rm det}(m_1, m_2, z)$ on a grid of values for subsequent interpolation. Following the procedure described in Refs. [36,37] and approximating detection as coming from a single interferometer, $p_{\rm det}$ is computed as

[1] We discuss the implications of recent BBH detections in the O3a run in Sec. VII.

[2] This point has also been recognized in the recent Ref. [34].

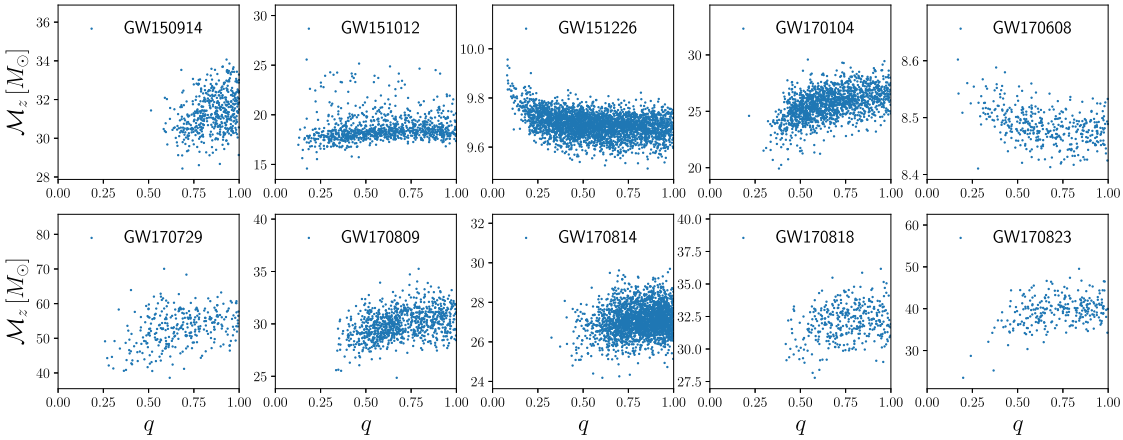

FIG. 1. Posterior samples from each of the sources in the O1O2 source catalog, indicated in each panel. We show samples in the plane of redshifted chirp mass $\mathcal{M}_z$ and mass ratio q, marginalized over the other source parameters. Samples have been thinned by a factor of 16 for visual clarity.

$$p_{\rm det}(m_1, m_2, z) = \int_{\rho_*/\rho_{\rm opt}(m_1,m_2,z)}^{1} p(\omega)\mathrm{d}\omega, \qquad (1)$$

where $\rho_{\rm opt}(m_1, m_2, z)$ is the signal-to-noise ratio (S/N) for an optimally oriented source, face on directly above the interferometer. The S/N threshold for detection is approximated as $\rho_* = 8$, ω encodes all the angular dependence of the interferometer response, and the orientation and angular position of the source have been marginalized over assuming isotropy, encoded in the distribution $p(\omega)$. The signal-to-noise ratio for an optimally oriented source is computed using the routines in the PyCBC software package [38]. To compute the noise power spectral density (PSD), we assume the IMRPHENOMD waveform approximant assuming nonspinning black holes,[3] and we approximate the PSD of each source in the GWTC-1 catalog with the PyCBC analytic function aLIGOEarlyHighSensitivityP1200087 [42]; i.e., we assume that each source is detected in a single aLIGO detector. This is a sufficiently good approximation for our purposes to the true PSD of each source, with the biggest difference arising at the lowest frequencies between $f = 10$ Hz and $f = 20$ Hz, a frequency range to which our final results are insensitive.

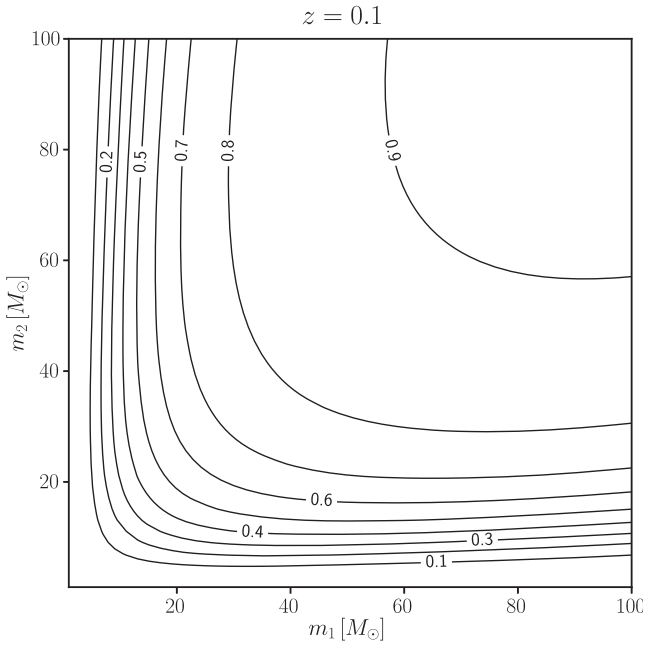

FIG. 2. Contours of constant detection probability at $z = 0.1$ as a function of the source-frame component masses, assuming the O1O2 observing specifications defined in the text.

In Fig. 2, we show contours of the detection probability for a source at $z = 0.1$ as a function of the source-frame component masses (see, e.g., Ref. [43] for similar plots). The S/N threshold results in a suppression of the detection probability below about 5 $M_\odot$. The figure makes it clear that LIGO in the O1O2 observing runs was sensitive to large mass ratios, with this increasingly true for the more sensitive O3 run, which has yielded objects with $q \approx 0.3$ [44] and $q \approx 0.1$ [45]. Note that the sensitivity falls to zero at very high (a few hundred solar) masses where the waveforms have no support above the minimum frequency $f_{\rm low}$.

In Fig. 3, we show the redshift dependence of $p_{\rm det}$ for a range of total source-frame masses and mass ratios. This dependence arises primarily from the d_L^{-1} dropoff in the S/N, but there is also a dependence via the redshifted (detector-frame) masses $(1+z)m_{1,2}$ at fixed source mass.

[3]Note that the assumption of nonspinning black holes should be a reasonable approximation in the case of PBH models, which are expected to have negligible spin at formation [39,40], although it is possible that subsequent accretion can lead to nonzero spin for the higher mass objects detected by LIGO [41]. In the case of astrophysical merger models, we will see that this is a reasonable approximation to the S/N.

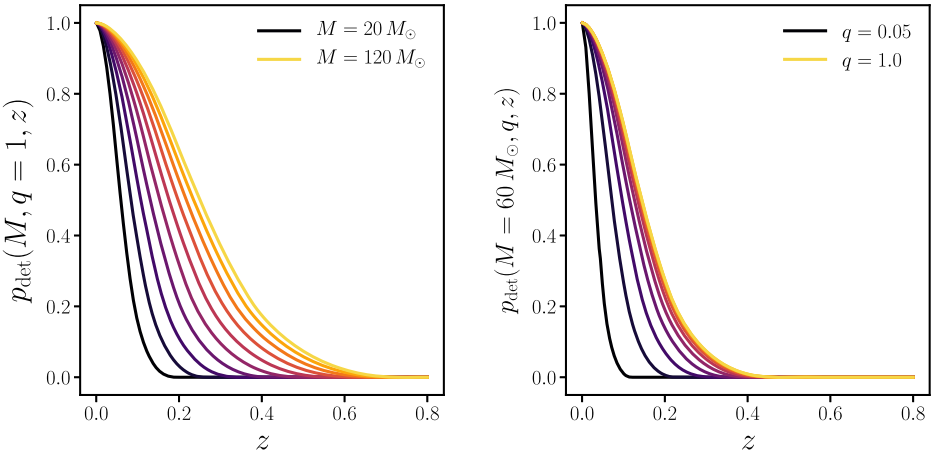

FIG. 3. Detection probability as a function of redshift for equal component masses and a range of total masses (left panel) and total mass $M = 60\ M_\odot$ and a range of mass ratios q (right panel). Warmer colors (increasing from bottom to top in both panels) indicate greater values of M or q.

Sources with redshifts $z \gtrsim 0.7$ are undetectable in the O1O2 runs for any component masses, with this upper limit quickly dropping as the total mass is lowered. For a total mass $M = 20\ M_\odot$, only sources with $z \lesssim 0.2$ are detectable, and then only for equal mass components. For comparison, the highest redshift in the GWTC-1 catalog is $z \approx 0.49$ (GW170729, total mass $M \approx 84\ M_\odot$, albeit with significantly nonzero spin), while the median redshift source is $z \approx 0.16$.

We close this section by noting that the methodology we have adopted for computing $p_{\rm det}$ is a commonly used approach in BBH population analyses (see, e.g., Refs. [29,30,37,46]). Nonetheless, as discussed in Appendix A of Ref. [31], our approximate method for computing the detection probability can overestimate the sensitive volume $\langle VT \rangle$ by up to a factor of 2 compared to the more accurate approach of injecting signals into the source detection pipeline (see their Fig. 9). This means that our constraints on overall merger rates are expected to be underestimated by at most a factor of 2. This translates into an underestimation of $f_{\rm PBH}$ of at most 20% for the models we consider.

III. BINARY MERGER RATE MODELS

In this section, we present the models of merging black hole binaries which we confront with the GWTC-1 catalog. The fundamental quantity we require for our statistical analysis is the differential merger rate density in the source frame as a function of masses and redshift.

A. PBH merger models

Many attempts have been made to model the merger rate of PBH binaries (e.g., Refs. [10,11,17,18,20,47,48]). As in Ref. [21], we use the formalism of Ref. [26], which itself builds upon that in Ref. [20]. We will first give a brief sketch of the calculation before presenting the resulting PBH merger rate. Readers unconcerned with the derivation may skip to Eqs. (8) and (9) where the differential merger rate is presented.

The calculation of Ref. [26] follows a pair of PBHs drawn from a mass function $\psi(m)$ (normalized to unity), initially comoving with the cosmic expansion in radiation domination. The PBH pair decouples from the expansion and forms a high-eccentricity binary, with gravitational torquing from other surrounding PBHs and a smoothly distributed dark matter component with small Gaussian fluctuations. The PBH binary forms with dimensionless angular momentum $j \ll 1$, emits GWs, and merges after a time [49],

$$\tau = \frac{3}{85}\frac{r_a^4}{\eta M^3} j^7, \tag{2}$$

where r_a is the semimajor axis of the binary, M is the total mass, and η is the symmetric mass ratio defined by $\eta = m_1 m_2/(m_1 + m_2)^2$. The ellipticity of the binary is given by $e = \sqrt{1 - j^2}$. The semimajor axis r_a follows from the dynamics of the system prior to binary formation and is given by $r_a \approx 0.1 a_{\rm dc} x_0$, where $a_{\rm dc}$ is the scale factor at decoupling and x_0 is an initial comoving separation. Decoupling takes place at roughly $a_{\rm dc} \approx a_{\rm eq}/\delta_b$, where $a_{\rm eq}$ is the scale factor at matter-radiation equality and $\delta_b \gg 1$ is the effective density fluctuation generated by the PBH pair, given by $\delta_b = (M/2)/\rho_M V(x_0)$ with ρ_M the background matter density and $V(x) = (4\pi/3)x^3$. We refer the reader to Ref. [26] for further discussion.

The time taken for a newly formed PBH binary to merge is a crucial factor in determining the rate of merging sources in the LIGO-Virgo sensitive volume, and Eq. (2) makes clear its high sensitivity to the angular momentum of the binary. At formation, this angular momentum is imparted

by gravitational torquing from other PBHs and fluctuations in the surrounding dark matter density. In the model of Refs. [20,26], these dark matter fluctuations are modeled as Gaussian with variance $\langle\delta_M^2\rangle$, with contributions from dark matter mass scales greater than approximately $10^{-3}M \approx 10^{-2}\ M_\odot$ for PBHs in the LIGO mass range. Following Refs. [20,50], we assume a fixed value $\langle\delta_M^2\rangle^{1/2} = 0.005$ extrapolating the linear adiabatic power spectrum measured on cosmic microwave background scales, although one should note that models of PBH formation typically invoke enhanced small-scale power or non-Gaussianity in the dark matter distribution, which could boost $\langle\delta_M^2\rangle$ significantly. The variance of angular momentum fluctuations in the vicinity of the PBH binary is then

$$\sigma_{j,M}^2 = \frac{6}{5} j_0^2 \frac{\sigma_M^2}{f_{\rm PBH}^2}, \tag{3}$$

where j_0 is a characteristic angular momentum (with $j_0 \ll 1$), $f_{\rm PBH} \equiv \rho_{\rm PBH}/\rho_{\rm DM}$ is the ratio of the PBH energy density to the dark matter energy density, and σ_M^2 is defined in Ref. [26] as a "rescaled variance" given by $\sigma_M^2 \equiv (\Omega_{\rm M}/\Omega_{\rm DM})^2\langle\delta_M^2\rangle$. Since the PBH binary is assumed to form deep in the radiation era when baryons are tightly coupled to photons, only the dark matter contributes to fluctuations in the local tidal field from the smooth matter component, and hence $\langle\delta_M^2\rangle$ should appear on the right-hand side of Eq. (3). The difference is negligible in comparison to the uncertainty on the variance on these small scales, however, so for consistency with Ref. [26], we take $\sigma_M \approx 0.006$.

The total variance of the angular momentum imparted to the PBH binary consists of the dark matter fluctuations plus those of the surrounding PBHs and is given by

$$\sigma_j^2 = \sigma_{j,M}^2 + \sigma_{j,\rm PBH}^2 = \frac{6}{5} j_0^2 \left(\frac{1+\sigma_m^2/\langle m\rangle^2}{\bar{N}(y)} + \frac{\sigma_M^2}{f_{\rm PBH}^2}\right), \tag{4}$$

where σ_m^2 is the variance of the PBH mass function, $\langle m\rangle$ is the average PBH mass (angle brackets denote expectation values over ψ/m), and $\bar{N}(y)$ is the expected number of PBHs within a comoving radius y of the binary. This latter quantity is needed because the model of Ref. [26] assumes that there is an *exclusion zone* around the binary of radius y, inside of which no other PBH can reside, lest its close proximity fatally disrupt the newly formed binary. The limit $\bar{N}(y) \to 0$ corresponds to no such exclusion.

The *distribution* of the angular momentum imparted to the binary follows from assuming Gaussianity for the dark matter and Poisson statistics for the surrounding PBHs. As shown in Ref. [20], this results in a Holtsmark distribution for the latter. The resulting probability density p_j for the angular momentum j is

$$jp_j(j) = \int_0^\infty {\rm d}u\, uJ_0(u) \exp\left[-\bar{N}(y)\int \frac{{\rm d}n(m)}{n} \times F\left(u\frac{m}{\langle m\rangle}\frac{1}{\bar{N}(y)}\frac{j_0}{j}\right) - u^2\frac{3}{10}\frac{\sigma_M^2}{f_{\rm PBH}^2}\frac{j_0^2}{j^2}\right], \tag{5}$$

where the innermost integral is over the number density of PBHs with

$$\frac{{\rm d}n}{{\rm d}m} = \rho_{\rm PBH}\frac{\psi(m)}{m}, \tag{6}$$

and $F(x) = {}_1F_2(-1/2; 3/4, 5/4; -9x^2/16) - 1$ with ${}_1F_2$ a generalized hypergeometric function. In the limit $\bar{N}(y) \to 0$ and $\sigma_M \ll f_{\rm PBH}$, we obtain the result of Ref. [20], and in the limit $\bar{N}(y) \to \infty$, we obtain a Rayleigh distribution for j with width σ_j, i.e., torquing only by the Gaussian dark matter fluctuations.

As in Refs. [14,21,26], we take

$$\bar{N}(y) = \frac{M}{\langle m\rangle}\frac{f_{\rm PBH}}{f_{\rm PBH}+\sigma_M}, \tag{7}$$

which agrees well with the numerical simulations of Ref. [26] for $f_{\rm PBH} \lesssim 10^{-1}$. To understand the form of this expression, first note the limiting case of a monochromatic mass function and $\sigma_M \ll f_{\rm PBH}$. In this case, $\bar{N}(y) = 2$; i.e., we expect two PBHs in the vicinity of the binary—the two components black holes themselves. This agrees with the discussion in Ref. [51] that y should be roughly the interparticle distance of the PBH distribution. In the case in which $\sigma_M \gg f_{\rm PBH}$, we have $\bar{N}(y) \to 0$; i.e., the exclusion region around the binary is expected to hold very few PBHs. Equation (7) extends this simple picture to the case of a broad mass function, identifying the transition value of $f_{\rm PBH}$ as $\sim\sigma_M$. We note that there is considerable uncertainty in the potential rate of disruption of newly formed binaries by surrounding PBHs, and deviations from Eqs. (5) and (7) may be expected in the case of broad mass functions.

The merger rate density at time τ of binaries in the model of Ref. [26] is given by ${\rm d}R = S \times {\rm d}R_0$, where

$$\begin{aligned}{\rm d}R_0 = {}& \frac{1.6\times10^6}{{\rm Gpc}^3\ {\rm yr}} f_{\rm PBH}^{\frac{53}{37}}\eta^{-\frac{34}{37}}\left(\frac{M}{M_\odot}\right)^{-\frac{32}{37}}\left(\frac{\tau}{\tau_0}\right)^{-\frac{34}{37}} \\ & \times\psi(m_1)\psi(m_2){\rm d}m_1{\rm d}m_2\end{aligned} \tag{8}$$

is the rate in the limit $\bar{N}(y) \to 0$ and $\sigma_M/f_{\rm PBH} \to 0$, with $\tau_0 = 13.8\times10^9$ yr, and S is the *suppression factor* given by

$$S = \frac{e^{-\bar{N}(y)}}{\Gamma(21/37)} \int_0^\infty \mathrm{d}v\, v^{-\frac{16}{37}} \exp\left[-\bar{N}(y)\langle m\rangle \int_0^\infty \frac{\mathrm{d}m}{m} \times \psi(m) F\left(\frac{m}{\langle m\rangle}\frac{v}{\bar{N}(y)}\right) - \frac{3\sigma_M^2 v^2}{10 f_{\mathrm{PBH}}^2}\right]. \quad (9)$$

The suppression factor S quantifies the effect of demanding that no PBHs be present in a region of size y around the binary [which would be expected to contain $\bar{N}(y)$ PBHs] *and* the effect of dark matter density fluctuations imparting angular momentum to the binary. It is straightforward to show that $0 \leq S \leq 1$.

We will often present results for a PBH model having no suppression factor, i.e., $S = 1$, where the merger rate is given by Eq. (8). Some of the uncertainty in the precise formation mechanism is bracketed by the cases $S = 1$ and the full expressions of Eqs. (8) and (9). More precise numerical simulations of PBH binary formation and evolution will be required for a more quantitative investigation of the sensitivity of our results to the formation model [15,52–55].

It is important to note that the PBH mass functions we use are based on calculations of the primordial mass function ψ. It is possible that ψ could evolve through mergers and/or accretion. So-called second-generation mergers, i.e., those involving one or more BHs which have already undergone a previous merger, are expected to be very rare compared with primary mergers [19,41] (although also see Ref. [56]). Accretion is a highly nonlinear process which is hard to model, but recently Ref. [41] has suggested this could play an important role on more massive PBHs due to the dark matter halo which forms around them at early times, and thereby acts as a significant additional gravitational attraction to nearby baryons (note that this only applies if $f_{\mathrm{PBH}} \ll 1$, but that is the case we consider in this paper). Reference [48] showed that the DM halo which forms around PBHs [52,57] has minimal impact on the merger rate and the estimate of f_{PBH}. Accretion onto a PBH can increase the initially negligible spin of a PBH, provided that the PBH mass grows significantly [41]. The most massive BH pair detected by LIGO is also the system with nonzero spin detected at highest significance, which may be consistent with a PBH model that includes a modest amount of accretion. However, the second-lightest BH merger event also has significant evidence for nonzero spin, suggesting that not all events are primordial. We will return to this issue in Sec. VII. We note that there exists a window for which accretion has a non-negligible impact on the spin of the most massive PBHs but has very little impact on the PBH mass function [14], but the impact of accretion is a highly nonlinear process which deserves further study; see e.g., Ref. [58].

Recently, Jedamzik has argued that the LIGO and Virgo results are consistent with $f_{\mathrm{PBH}} = 1$ provided that PBHs follow a broad mass function with a large spike at 1 $M_\odot$ [15,53], as motivated by the softening of the equation-of-state parameter during the Standard Model QCD phase transition [23,59]. This result is based in numerical simulations of dense PBH clusters, and it does not appear to be in contradiction with the constraints we derive of $f_{\mathrm{PBH}} \ll 1$ based on a relatively narrow mass function; see, e.g., Refs. [24,26], which show that the analytic estimate for the merger rate which we used may be unreliable if $f_{\mathrm{PBH}} \gtrsim 0.1$. See also Refs. [52,54,55] for further numerical studies of the PBH binary disruption rate.

Finally, our baseline results assume a log-normal mass function for the PBHs, given by

$$\psi(m) = \frac{1}{m\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\ln^2(m/m_c)}{2\sigma^2}\right], \quad (10)$$

where m_c is the peak of the function $m\psi(m)$ and σ is its logarithmic width. This is a good approximation to the PBH mass function in the case of formation from a smooth, symmetric peak in the power spectrum [60,61], although deviations are expected in the case of particularly narrow power spectrum peaks [62]. We consider the case of non-log-normal mass functions in Sec. VI.

One consideration when using the model of Ref. [26] for log-normal mass functions is the oversuppression of the merger rate for very broad mass functions. Physically, one would expect a population of very light PBHs to have little effect on the merger rate in the LIGO mass range, since this light population does not contribute significantly to the gravitational attraction between two heavier PBHs, leaving negligible impact on the formation of the binary. However, in the calculation of Ref. [26], a large population of light black holes makes a large contribution to the expected number of PBHs in the vicinity of the binary, $\bar{N}(y)$, and is assumed to cause disruption to the heavier pair of PBHs [21]. To ensure that this oversuppression does not affect the constraint presented in this paper, we quantify the value of σ for which the log-normal distribution becomes broad enough that the suppression becomes significant. We do this by considering the differential merger rate for an equal mass merger in two cases: considering the full mass function and a mass function with a low-mass cutoff, so that the population of light black holes is removed. For three different values of the PBH mass m (the same for both PBHs in the binary), the inclusion of the low-mass population causes a significant suppression (greater than 10 orders of magnitude) for $\sigma \gtrsim 2$. As we shall see later, this is well above the range that the data favor, and so we assume that the model in Ref. [26] is valid for the constraints we present. We note that a thorough investigation of this effect will require the running of N-body simulations.

To compute the merger rate as a function of mass and redshift, we use a fast and accurate approximation to the

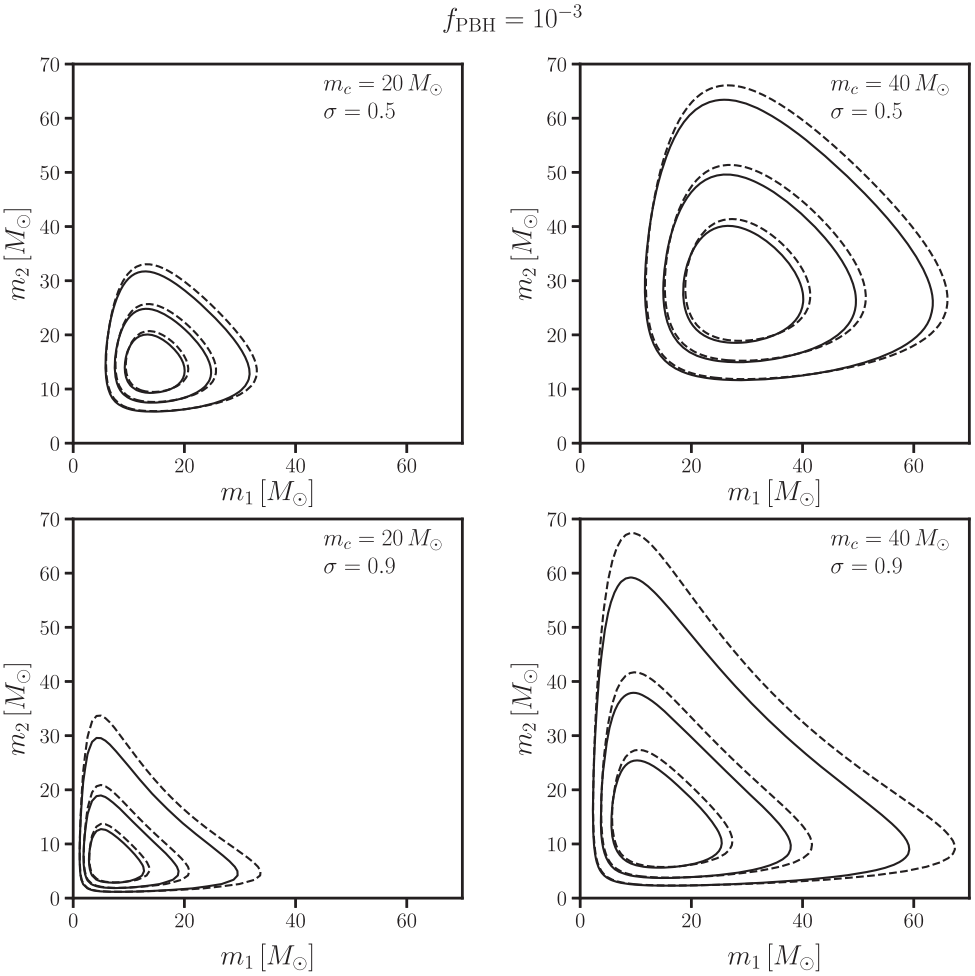

FIG. 4. Source-frame differential merger rate density $\mathrm{d}R/\mathrm{d}m_1\mathrm{d}m_2$ for log-normal PBH mass functions at $z = 0.1$ and $f_{\rm PBH} = 10^{-3}$, as a function of the individual source-frame masses. Solid and dashed contours are predictions with and without the suppression factor accounting for three-body interactions, respectively, and contours are drawn at 25%, 50% and 75% of the peak value. The log-normal mass-function parameters are indicated in the top-right corner of each panel.

suppression factor valid for log-normal mass functions, described in Appendix A.

In Fig. 4, we show the dependence of the source-frame differential merger rate density $\mathrm{d}R/\mathrm{d}m_1\mathrm{d}m_2$ on the component masses, for several choices of the log-normal mass-function parameters. The log-normal distribution has a characteristic skewness toward large masses, giving rise to a skewness toward large total mass. This also gives rise to a broad range of mass ratios, as seen by the off-diagonal extent of the merger rates, which increases with σ.

Importantly, for $\sigma \lesssim 1$, the shape of the merger rate distribution is primarily controlled by the mass function terms $\psi(m_1)\psi(m_2)$, with only limited sensitivity to the factors of M and η which multiply these terms in Eq. (8). These factors primarily control the shape of the tails of the distribution. When $\sigma \gtrsim 1$, the mass function is broad, and $\psi(m_1)\psi(m_2)$ varies more slowly over a fixed mass range, such that factors of M and η can be relatively more important.

This behavior is also seen in the importance of the suppression factor S, indicated by the difference between the solid and dashed lines in Fig. 4. The influence of S on the shape of the merger rate distribution is weak for the mass function parameters plotted (which we will see correspond to those favored by the data), increasing in importance for larger σ. The suppression factor depends only on the total mass, via Eq. (7), with this dependence weakening for small $f_{\rm PBH}/\sigma_M$.

Note also that the peak of the log-normal mass function is at $m_c e^{-\sigma^2}$, its mean is at $m_c e^{\sigma^2/2}$, and its median is at m_c; i.e., for large σ, the distribution is significantly skewed.

The rate of merger events observable today can be found by integrating $\mathrm{d}R$ over mass and volume and is given by

$$\frac{\beta}{T_{\rm obs}} = \int \mathrm{d}z\,\mathrm{d}m_1\mathrm{d}m_2 \frac{1}{(1+z)}\frac{\mathrm{d}V_c}{\mathrm{d}z}\frac{\mathrm{d}R}{\mathrm{d}m_1\mathrm{d}m_2}\,p_{\rm det}(m_1, m_2, z), \tag{11}$$

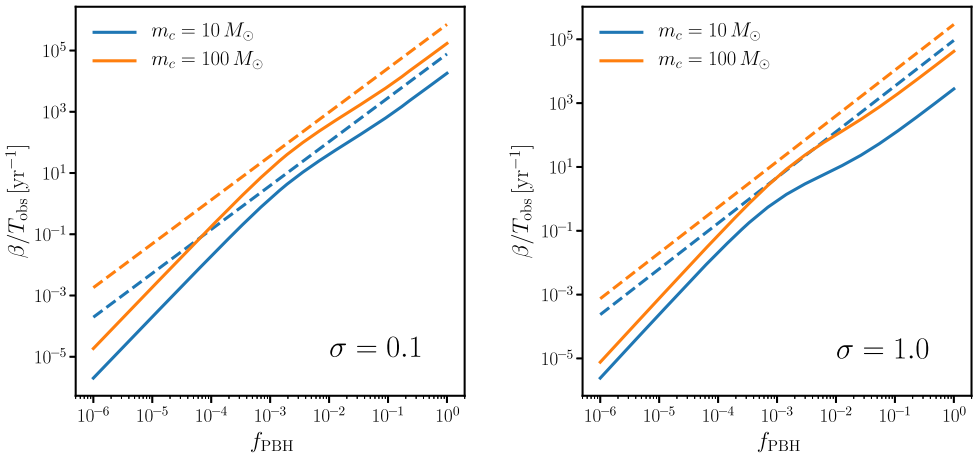

FIG. 5. Rate of detectable PBH mergers, per year as a function of $f_{\rm PBH}$ with (solid) and without (dashed) the suppression factor, for $m_c = 10\ M_\odot$ (blue) and $m_c = 100\ M_\odot$ (orange), and for $\sigma = 0.1$ (left panel) and $\sigma = 1.0$ (right panel).

where $dR = S \times dR_0$, with dR_0 and S given by Eqs. (8) and (9), and dV_c is the comoving volume of a thin spherical shell of width dz. The factor of $(1+z)^{-1}$ accounts for the difference between proper (source-frame) rate and observed (detector-frame) rate. The total number of detectable mergers is β, and $T_{\rm obs}$ is the observation time. Note that Eq. (11) assumes that $p_{\rm det}$, and hence the interferometer PSD, is independent of time. We make this approximation throughout our analysis.

In Fig. 5, we plot the detectable merger rate $\beta/T_{\rm obs}$ as a function of $f_{\rm PBH}$, for representative values of m_c and σ, and using the $p_{\rm det}$ described in Sec. II. For a six-month observation period, the figure shows that ten binary merger events would be expected for $f_{\rm PBH} \sim$ a few $\times\, 10^{-3}$, with $f_{\rm PBH} \sim 10^{-2}$ possible for large σ and $m_c \sim 10\ M_\odot$. Note that $f_{\rm PBH}$ primarily controls the *amplitude* of the merger rate, with only a minor impact on its mass dependence. This will be important when testing these models against the LIGO data.

B. LIGO empirical merger models

As well as PBH models of binary mergers, we also consider two empirical distributions often used to model BBH populations. These models, termed model A and model B, were introduced in Refs. [27,63] and extended in Refs. [28,43,64,65]. We use the forms as presented in Ref. [31].

Both models can be described by an intrinsic merger rate given by

$$\frac{dR}{dm_1 dm_2} = \begin{cases} R_0 C(m_1) m_1^{-\alpha} q^{\beta_q} & \text{if } m_{\rm min} \le m_2 \le m_1 \le m_{\rm max} \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

where R_0 is a constant amplitude, $q = m_2/m_1$, and $C(m_1)$ is such that the marginal distribution for the heavier mass is $m_1^{-\alpha}$. Model A fixes $m_{\rm min} = 5\ M_\odot$ and $\beta_q = 0$, allowing $m_{\rm max}$, R_0, and α to vary. Model B allows all five parameters to vary. When $\beta_q = 0$, we have $C(m_1) \propto 1/(m_1 - m_{\rm min})$. In the default formulation of these models, the rate R_0 is assumed to be independent of redshift.

The two models considered here are not intended to be detailed physical models of the merger rate of stellar black holes; rather, they are empirical parametrizations that allow for straightforward computation for comparison with data. They do, however, have two important features, motivated by astrophysics, which will prove crucial when comparing with PBH models; the upper and lower cutoff in mass. The lower mass cutoff is motivated by observations of x-ray binaries [66] and appears to be roughly 5 $M_\odot$, with a mass gap expected between this and the predicted upper limit for a neutron star of roughly 2 $M_\odot$ (see, however, Ref. [45] for a recent detection of a compact object in this mass gap). The upper limit is more uncertain and is partly motivated *a posteriori* from the GWTC-1 catalog. There is some astrophysical motivation for an upper limit of roughly 50 $M_\odot$ from pulsational pair-instability supernovae and potentially a mass gap between 50 and 150 $M_\odot$ due to the combined effect of this with pair instability (see Ref. [16] for relevant references). Importantly, the PBH models do not require upper or lower limits for the black hole mass, which will prove crucial for discriminating models. The detailed power-law behavior of models A and B is less physically motivated, so the LIGO models may be seen as a combination of astrophysical and empirical considerations.

The models considered in Ref. [31] also specify distributions for the spin parameters [67]. To facilitate model comparison with the PBH merger scenario, where there is considerable uncertainty in the form of the spin distribution, we will neglect information from spin by marginalizing over it as discussed further in Sec. IV. We note in passing that all but two of the sources (GW151226 and GW170729) in the GWTC-1 catalog are consistent with

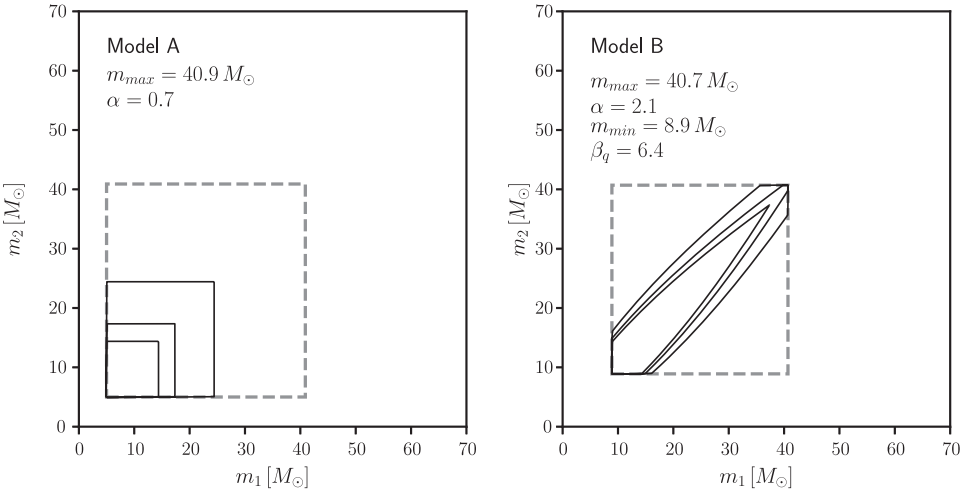

FIG. 6. Source-frame differential detectable merger rate for LIGO model A (left panel) and model B (right panel), as a function of the individual source-frame masses. The population parameters are indicated in the top left of each panel and correspond to the values maximizing the posterior of the O1O2 data. Contours are drawn at 25%, 50%, and 75% of the peak value, and the dashed lines indicate the boundaries of the distribution where the rates go to zero.

zero spin at 90% confidence. We refer the reader to Ref. [32,68] for recent Bayesian analyses using spin in the PBH context and Ref. [69] for a Bayesian study including a zero-spin black hole population, which could be considered a simplified proxy for a PBH model.

In Fig. 6, we show the mass dependence of the merger rate for models A and B, analogous to Fig. 4 in the case of PBHs, for representative values of the models' parameters. Note that for consistency with the PBH models we have extended the definition of m_1 and m_2 in Eq. (12) to the full mass plane, removing the requirement that $m_2 \leq m_1$. In the case of model A, the contours are perfect squares, being independent of the lighter mass for fixed heavier mass. For model B with positive β_q, the merger rate is strongest for $q \approx 1$, and the contours are hence more concentrated around the diagonal. That the distributions in Fig. 6 appear somewhat unphysical is a consequence of the explicit symmetry breaking between m_1 and m_2 in Eq. (12). We emphasize that these distributions are not intended to be realistic models of astrophysical black hole binary formation but rather capture broad features in the source population distributions.

A more complex model, dubbed model C, is also analysed in Ref. [31] and found to be superior fit to the GWTC-1 sources, although it is not significantly preferred over model B. Both model B and model C were found to be better fits than model A, so for simplicity, we only consider the two models A and B, noting that Bayesian evidence ratios against model C can be easily deduced from Table 3 or Ref. [31]. A future extension of this work would be to compare PBH source distributions with more realistic models incorporating relevant astrophysical effects, in the manner of Refs. [37,65].

IV. STATISTICAL FRAMEWORK

To assess the merger rate models described in Sec. III against data from the BBH sources in the GWTC-1 catalog, we must specify a likelihood function. As described in, e.g., Refs. [29,70,71], the likelihood of the observed data $\mathbf{d} = \{\mathbf{d}_i\}_{i=1\ldots N_{\rm obs}}$ from $N_{\rm obs}$ sources given population hyperparameters $\boldsymbol{\theta}$ and a merger rate model M is

$$p(N_{\rm obs}, \mathbf{d}|\boldsymbol{\theta}, M) \propto \left[\prod_{i=1}^{N_{\rm obs}} \int {\rm d}m_1 {\rm d}m_2 {\rm d}z\, p(\mathbf{d}_i|m_1, m_2, z) \times \frac{{\rm d}N}{{\rm d}m_1 {\rm d}m_2 {\rm d}z}(m_1, m_2, z|\boldsymbol{\theta}, M)\right] e^{-\beta(\boldsymbol{\theta}, M)}, \tag{13}$$

where $p(\mathbf{d}_i|m_1, m_2, z)$ is the likelihood of each observed dataset (i.e., GW strain data) given the source masses and redshift (marginalizing over all other parameters), dN is the number of mergers in a mass and redshift interval, and $\beta(\boldsymbol{\theta}, M)$ is defined in Eq. (11), i.e.,

$$\frac{{\rm d}N}{{\rm d}m_1 {\rm d}m_2 {\rm d}z}(m_1, m_2, z|\boldsymbol{\theta}, M) = T_{\rm obs} \frac{1}{(1+z)} \frac{{\rm d}V_c}{{\rm d}z} \frac{{\rm d}R}{{\rm d}m_1 {\rm d}m_2}(m_1, m_2, z|\boldsymbol{\theta}, M). \tag{14}$$

This likelihood correctly accounts for the interferometer selection function via $p_{\rm det}$, which enters via the expected number of detected mergers β, and the uncertainties on the parameters of each source, which may be correlated. It assumes each source is independent. Since our data cover the O1 and O2 observing runs of LIGO, we take $T_{\rm obs} =$ 169.5 days.

As shown in Ref. [71], the likelihood in Eq. (13) is the product of a Poisson likelihood for the observed number of detections $N_{\rm obs}$ when $\beta(\boldsymbol{\theta}, M)$ were expected with a likelihood for observable data $\mathbf{d}_i$, the result being an inhomogeneous Poisson likelihood. Note that the selection function $p_{\rm det}$ only enters via β, since the observed data are observable by definition. The likelihood thus consistently accounts for information on population models coming from the observed distribution of source parameters and the overall number of detections—these were considered separately for PBH models in a frequentist approach in Ref. [21].

We note that several works constraining PBH merger models with GW data have used likelihoods differing from Eq. (13), for example, Refs. [14,26]. We emphasize that Eq. (13) correctly accounts for the source parameter correlations and selection effects, and its use ensures that posteriors on the PBH model parameters are unbiased.[4]

We approximate the integral over the source likelihoods in Eq. (13) with a sum over Monte Carlo samples from the source posteriors available from the GWTC-1 catalog [16]. To do this, we need to first divide out the source prior which was used in the LIGO inference; i.e., we have

$$p(N_{\rm obs}, \mathbf{d}|\boldsymbol{\theta}, M) \propto \left[\prod_{i=1}^{N_{\rm obs}} \left\langle \frac{1}{\pi(z^i, m_1^i, m_2^i)} \frac{\mathrm{d}N}{\mathrm{d}m_1 \mathrm{d}m_2 \mathrm{d}z}(m_1^i, m_2^i, z^i|\boldsymbol{\theta}, M) \right\rangle_{\{z^i, m_1^i, m_2^i\}}\right] e^{-\beta(\boldsymbol{\theta}, M)}, \tag{15}$$

where $\pi(z^i, m_1^i, m_2^i)$ is the prior on parameters for source i and angle brackets denote an expectation value over Markov-chain Monte Carlo (MCMC) samples from the source posterior. The prior is uniform in the detector-frame masses and scales as d_L^2 in the space of luminosity distance d_L, i.e.,

$$\pi(z, m_1, m_2) \propto (1+z)^2 d_L^2(z) \frac{\mathrm{d}d_L(z)}{\mathrm{d}z}. \tag{16}$$

We ignore any information coming from the spin of the black holes by marginalizing over spin parameters when averaging over the MCMC samples—this is equivalent to assuming that the merger rates are independent of spin. Similarly, we assume all intrinsic merger rates are independent of angular position and orientation and marginalize over these parameters. This is also implemented in our treatment of the detection probability $p_{\rm det}$ in Eq. (1), which makes the implicit assumption that the signal-to-noise ratio is not significantly impacted by the component spins.[5]

We use Bayes's theorem to compute the posterior of the population hyperparameters, $p(\boldsymbol{\theta}|\mathbf{d}, N_{\rm obs}, M) \propto p(\boldsymbol{\theta}|M) p(\mathbf{d}, N_{\rm obs}|\boldsymbol{\theta}, M)$. For our log-normal PBH mass function, we have $\boldsymbol{\theta} = [f_{\rm PBH}, m_c, \sigma]$. This requires us to specify priors on the population hyperparameters. As is common in Bayesian inference problems, the choice of these priors is somewhat arbitrary. We will see that the data are sufficiently constraining that the priors have negligible impact on posterior parameter constraints but can significantly impact Bayesian evidences. We will see later that this latter prior dependence can be unpicked using a suitable approximation to the evidence.

In Table I, we show the priors adopted in our inference runs. In the case of the LIGO models A and B, we use the priors adopted in Ref. [31]. In the case of the PBH models, we take a uniform prior on $\log f_{\rm PBH}$ motivated by the fact that $f_{\rm PBH}$ primarily controls the amplitude of the merger rate and its order of magnitude is unknown. We assume uniform priors on $\log m_c$ and $\log \sigma$ since their orders of magnitude are similarly unconstrained *a priori* and for the reason that these would be the Jeffreys priors on these parameters if the likelihood were proportional to $\psi(m)$.[6]

Finally, we assume that all the BBH sources in the catalog are primordial in origin when performing inference under a PBH model and that all sources are astrophysical when using model A or model B. In principle, we should account for the possibility that some binaries consist of PBH pairs and some are astrophysical pairs (the merger rate of mixed PBH-astrophysical black hole binaries [72,73] is expected to be small compared with that of PBH-PBH binaries for the values of $f_{\rm PBH}$ we consider). This could be implemented by introducing an extra parameter controlling the proportion of sources in each formation channel [32,74]. For simplicity, we do not take this approach, and instead treat all ten sources as either primordial or

[4]Our approach is more similar to the recent Ref. [19], differing in our self-consistent treatment of $f_{\rm PBH}$ in the merger rate amplitude and our use of a suppression factor accounting for three-body effects.

[5]Alternatively, our neglect of spin can be phrased as the imposition of zero spin in all source components plus the assumption that the spin parameters are uncorrelated with the inferred masses and redshifts. This latter assumption allows us to include all the posterior samples when computing Eq. (15) and not just those lying in the zero-spin hypersurface. This is a reasonable approximation for the GWTC-1 sources.

[6]Note that the Jeffreys prior for a Poisson distribution with rate parameter λ is $\pi(\lambda) \propto \lambda^{-\frac{1}{2}}$, i.e., uniform in $\sqrt{\lambda}$. In the PBH model, the amplitude of the merger rate scales roughly as $f_{\rm PBH}^{\frac{53}{37}}$, so an uninformative prior might be expected to scale roughly as $f_{\rm PBH}^{-\frac{53}{74}} \approx f_{\rm PBH}^{-0.7}$, i.e., uniform in $f_{\rm PBH}^{\frac{21}{74}}$. Our prior, scaling like $f_{\rm PBH}^{-1}$, is therefore approximately uninformative.

TABLE I. Priors used in this work. All priors are uniform within the limits given in the right-hand column. Models sharing parameters which vary have the same priors on those parameters. See the main text for the definition of these parameters.

Parameter	Prior
$\log_{10} f_{\rm PBH}$	$[-6, 0]$
$\log_{10} m_c(M_\odot)$	[0, 4]
$\log_{10} \sigma$	$[-1, 0.7]$
$\log_{10} R_0$	$[-1, 3]$
$m_{\rm max}(M_\odot)$	[30, 100]
$m_{\rm min}(M_\odot)$	[5, 10]
α	$[-4, 12]$
β_q	$[-4, 12]$
λ	[0, 0.5]
$\log_{10} m_{c,1}(M_\odot)$	$[-1, 3]$
$\log_{10} m_{c,2}(M_\odot)$	$[-1, 3]$

astrophysical, using Bayesian model selection to compare how well the respective models fit the data. Our constraints on $f_{\rm PBH}$ should thus be interpreted as upper limits.

V. BAYESIAN INFERENCE FROM THE GWTC-1 CATALOG

We use the likelihood in Eq. (15) with the models described in Sec. III and priors listed in Table I to draw samples from the posterior distribution of each model's parameters. We use the nested sampling algorithm [75,76] with multiellipsoidal bounded sampling [77] as implemented in DYNESTY [78] to draw samples from the posterior.

The nested sampling algorithm also computes the evidence for each model M, integrating over the prior as

$$p(M|\mathbf{d}, N_{\rm obs}) \propto p(M) \int \mathrm{d}\boldsymbol{\theta}\, p(\mathbf{d}, N_{\rm obs}|\boldsymbol{\theta}, M) p(\boldsymbol{\theta}|M), \quad (17)$$

where $p(M)$ is the prior on the model. We assume that $p(M)$ is uniform, such that Bayes factors are equivalent to evidence ratios

$$\frac{Z_{M_1}}{Z_{M_2}} \equiv \frac{\int \mathrm{d}\boldsymbol{\theta}\, p(\mathbf{d}, N_{\rm obs}|\boldsymbol{\theta}, M_1) p(\boldsymbol{\theta}|M_1)}{\int \mathrm{d}\boldsymbol{\theta}\, p(\mathbf{d}, N_{\rm obs}|\boldsymbol{\theta}, M_2) p(\boldsymbol{\theta}|M_2)}. \quad (18)$$

We express all evidences relative to that of model B and quote errors on the evidence using the default first-order approximation produced by DYNESTY.[7]

In Table II, we present the marginalized parameter constraints on the parameters of each model, as well as the evidence relative to model B. We will first discuss the parameter constraints on the baseline PBH models and the empirical LIGO models A and B, before discussing the evidences and model consistency tests. We will then introduce the extensions to the baseline models, the results of which are also listed in Table II for completeness.

A. Parameter constraints

1. PBH models

Our baseline PBH model uses the merger rate model of Eq. (8) with a suppression factor given in Eq. (9). To study the influence of the suppression factor, we also consider a model with $S = 1$. Differences between these two models can be roughly interpreted as encapsulating the uncertainty associated with PBH binary disruption.

In Fig. 7, we plot two-dimensional Bayesian credibility intervals (68% and 95% weighted posterior quantiles) for $\log_{10} f_{\rm PBH}$ and the log-normal mass function parameters $\log_{10} m_c$ and $\log_{10} \sigma$, along with the marginalized one-dimensional posteriors. Note that the priors on these parameters are uniform, with limits given in Table I. Constraints on these, and the derived parameters m_c and σ, are given in Table II.

The posterior constraints (median and 95% credible intervals) on these parameters are

$$\begin{aligned} \log_{10} f_{\rm PBH} &= -2.30^{+1.16}_{-0.35} \\ m_c &= 24.23^{+528.62}_{-6.31} M_\odot \\ \sigma &= 0.82^{+1.71}_{-0.35}. \end{aligned} \quad (19)$$

Figure 7 demonstrates that the posterior is highly non-Gaussian, with a pronounced curving degeneracy between all three parameters. There is, however, a clear peak around the median values quoted above, with a preferred value of $f_{\rm PBH} \approx 0.005$, assuming all the BBHs in GWTC-1 are primordial. As hinted at in Sec. III, $f_{\rm PBH} = 1$ is strongly disfavored, with this model drastically overproducing BBH mergers. The preferred mass function parameters roughly correspond to the average mass of the components in the catalog and the approximate spread in values.

The degeneracy tail in Fig. 7 skews the one-dimensional posteriors to large values of m_c, σ, and $f_{\rm PBH}$. This tail (also visible in the likelihood plots in Ref. [26]) is a three-parameter degeneracy caused by the suppression factor, Eq. (9). We investigate its origin in detail in Appendix B. Briefly, the suppression factor can allow for enhanced $f_{\rm PBH}$ without overproducing mergers by increasing $\bar{N}(y)$, since $S \approx e^{-\bar{N}(y)}$. We compute this using Eq. (7), which depends on the log-normal mass function parameters as $\bar{N}(y) \propto M/\langle m \rangle = (M/m_c) e^{\sigma^2/2}$. Large-$\sigma$ mass functions are highly skewed; the total mass M is typically approximately $2m_c$, meaning a high proportion of masses in the integral contributing to β, Eq. (11), have $M \gg \langle m \rangle$ when $\sigma \gtrsim 1$, giving large suppression factors. Models with high σ and

[7]We verify that this approximation to the evidence agrees with the more accurate `simulate_run` approximation in DYNESTY to within 10% in all cases and also agrees well with resampled and jittered approximations to the evidence.

TABLE II. Median and 95% credible intervals for the parameters of each model considered. The bottom four rows display difference in best-fit log likelihood between each model and LIGO model B, the log of the Occam factor defined in the text, the difference in log evidence between the DYNESTY nested sampling estimate and the Laplace approximation defined in the text, and the Bayesian evidence ratios computed from nested sampling along with uncertainties.

	Model					
Parameter	PBH	PBH, $S=1$	Model A	Model B	PBH, $S=1$, $m_{\max}=50\ M_\odot$	PBH, $S=1$, skew-bimodal
$\log_{10} f_{\rm PBH}$	$-2.30^{+1.16}_{-0.35}$	$-2.76^{+0.25}_{-0.24}$	...	...	$-2.72^{+0.25}_{-0.25}$	$-2.74^{+0.23}_{-0.23}$
$\log_{10} m_c(M_\odot)$	$1.38^{+1.36}_{-0.13}$	$1.26^{+0.12}_{-0.22}$	...	...	$1.91^{+1.91}_{-0.76}$	...
$\log_{10} \sigma$	$-0.09^{+0.49}_{-0.24}$	$-0.21^{+0.24}_{-0.16}$	...	...	$0.27^{+0.23}_{-0.47}$	...
$m_c(M_\odot)$	$24.23^{+528.62}_{-6.31}$	$18.06^{+5.72}_{-7.10}$	...	...	$81.28^{+6525.7}_{-67.15}$	...
σ	$0.82^{+1.71}_{-0.35}$	$0.61^{+0.45}_{-0.19}$	...	...	$1.86^{+1.30}_{-1.23}$	...
$\log_{10} R_0$	...	...	$1.63^{+0.50}_{-0.45}$	$1.55^{+0.41}_{-0.43}$	...	...
$m_{\max}(M_\odot)$	...	...	$42.65^{+18.96}_{-5.99}$	$42.73^{+35.11}_{-6.31}$	50.0	...
$m_{\min}(M_\odot)$	...	...	5.00	$7.88^{+1.30}_{-2.64}$	...	...
α	...	...	$0.94^{+1.59}_{-2.38}$	$1.93^{+1.70}_{-1.96}$	...	...
β_q	...	...	0.00	$6.62^{+5.04}_{-6.62}$	...	...
λ	...	...	...	...	...	$0.35^{+0.14}_{-0.27}$
$\log_{10} m_{c,1}(M_\odot)$	...	...	...	...	...	$1.08^{+0.57}_{-0.38}$
$\log_{10} m_{c,2}(M_\odot)$	...	...	...	...	...	$1.57^{+0.08}_{-0.62}$
$m_{c,1}(M_\odot)$	...	...	...	...	...	$12.02^{+32.65}_{-10.82}$
$m_{c,2}(M_\odot)$	...	...	...	...	...	$37.15^{+7.52}_{-28.24}$
$\ln L^*/L^*_{\rm B}$	−6.99	−7.14	−2.51	0.00	−5.44	−3.53
ln Occam	−6.13	−8.21	−5.71	−6.74	−5.46	−7.73
$\ln Z_{\rm Lap}/Z_{NS}$	$1.60^{+0.16}_{-0.16}$	$0.26^{+0.17}_{-0.17}$	$0.77^{+0.15}_{-0.15}$	$0.63^{+0.16}_{-0.16}$	$0.54^{+0.13}_{-0.13}$	$1.92^{+0.18}_{-0.18}$
$\ln Z_{NS}/Z_{NS,\rm B}$	$-7.35^{+0.23}_{-0.23}$	$-8.25^{+0.23}_{-0.23}$	$-1.62^{+0.22}_{-0.22}$	0.00	$-4.01^{+0.21}_{-0.21}$	$-5.79^{+0.24}_{-0.24}$

high $f_{\rm PBH}$ also need high m_c in order to give an acceptable fit to the approximately 10 solar mass region occupied by the LIGO sources. Fixing M to the LIGO mass scale implies that m_c must be increased when σ is increased to keep $m_c e^{-\sigma^2/2}$ fixed in order to keep the suppression factor constant in the observed mass range. This results in a three-parameter degeneracy allowing for $f_{\rm PBH}$ as high as 0.07. We note that this partly arises due to the ambiguity of defining a "typical" mass scale in models with highly skewed and broad mass functions, which raises concerns about the validity of Eq. (7). We note that the peak of the posterior is reasonably robust to the degeneracy tail and that more accurate simulations will be needed to investigate the formation and evolution of PBH binaries with these extreme mass functions.

This explanation for the degeneracy tail is supported by Fig. 8, which shows the posterior for the PBH model with the suppression factor set to unity. In this case, there is no mechanism available to suppress the merger rate when $f_{\rm PBH}$ is high, and the constraints are much more Gaussian and confined in parameter space.

The constraints on the parameters for this $S=1$ model are (median and 95% credible interval)

$$\begin{aligned} \log_{10} f_{\rm PBH} &= -2.76^{+0.25}_{-0.24} \\ m_c &= 18.06^{+5.72}_{-7.10} M_\odot \\ \sigma &= 0.61^{+0.45}_{-0.19}. \end{aligned} \tag{20}$$

These constraints arise from fitting the observed mass scale and spread in observed masses (which effectively fix m_c and σ) and fitting $f_{\rm PBH}$ to match the observed rate of mergers. There is a slight tendency for the data to simultaneously prefer low values of m_c and high values σ, as seen in the right-hand panel of Fig. 8. This combination keeps the merger rate roughly constant in the observed mass region, although the degeneracy is weak.

The median values of the parameters are fairly stable to switching on the suppression factor, with smaller values of m_c and σ now preferred due to the absence of the degeneracy tail and the median $f_{\rm PBH}$ now ≈ 0.0017, i.e., almost a factor of 3 smaller. Once again, $f_{\rm PBH} \approx 1$ is highly disfavored.

We note that for both PBH models the priors chosen are sufficiently broad that they do not influence the posterior constraints on the parameters, as shown by Fig. 7 and the left panel of Fig. 8.

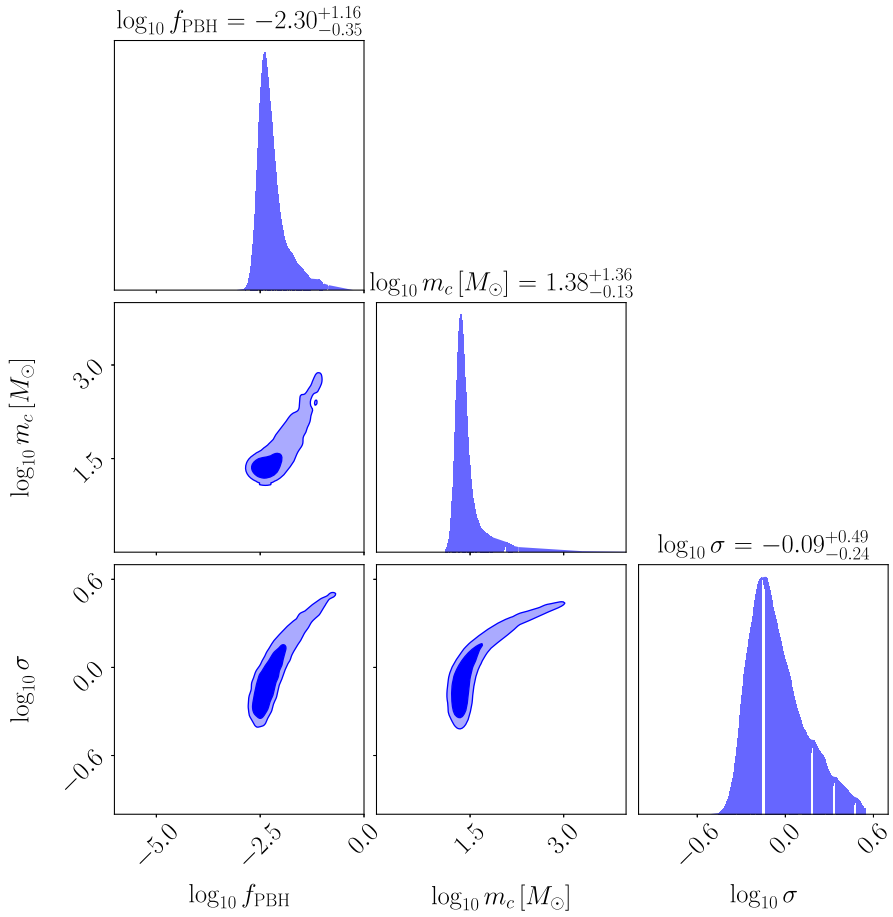

FIG. 7. *Off-diagonal panels*: Two-dimensional 68% and 95% marginal posterior quantiles for the parameters of the log-normal PBH model including the three-body suppression factor, given the GWTC-1 data. The plot boundaries correspond to the extent of the (uniform) priors on the parameters shown. *Diagonal panels*: One-dimensional marginal posterior densities for the parameters. Above each panel are the marginalized posterior median and 95% posterior quantiles for each parameter.

2. LIGO model A and model B

In Fig. 9, we show the posterior constraints on the parameters of the empirical LIGO models A and B, using the priors listed in Table I. These posteriors are fully consistent with those presented in Ref. [31], with only weak constraints provided on the model B parameters $m_{\rm min}$ and β_q. In contrast, the constraints on the upper component mass limit $m_{\rm max}$ and the power law slope in the distribution of the heavier mass α are reasonably well constrained, being determined by the maximum component mass in the catalog and the typical spread in masses, respectively (cf. the m_c and σ parameters of the PBH model). The amplitude of the merger rate simply fits the observed number of mergers, analogous to $f_{\rm PBH}$ in the PBH model.

The parameter constraints (median and 95% credible intervals) for model A are

$$\begin{aligned} m_{\rm max} &= 42.65^{+18.96}_{-5.99} M_\odot \\ \alpha &= 0.94^{+1.59}_{-2.38} \\ R_0 &= 42.66^{+92.24}_{-27.52}\ {\rm Gpc^{-3}\,yr^{-1}}, \end{aligned} \tag{21}$$

while those for model B are

$$\begin{aligned} m_{\rm max} &= 42.73^{+35.11}_{-6.31}\ M_\odot \\ \alpha &= 1.93^{+1.70}_{-1.96} \\ R_0 &= 35.48^{+55.72}_{-22.30}\ {\rm Gpc^{-3}\,yr^{-1}} \\ m_{\rm min} &= 7.88^{+1.30}_{-2.64}\ M_\odot \\ \beta_q &= 6.62^{+5.04}_{-6.62}. \end{aligned} \tag{22}$$

These constraints are consistent with those presented in Ref. [31], with the exception of R_0, which we find to be typically smaller with $R_0^{\rm LIGO}/R_0^{\rm here} \approx 1.50$ for both models. This can be explained by the difference in $p_{\rm det}$ arising from using the semianalytic approximation described in Sec. II vs a more accurate method using pipeline injections, as discussed in Appendix A of Ref. [31]. Our approximation overestimates the LIGO sensitive volume by between a factor of 1.4 and 1.9 depending on $m_{\rm max}$ and α (top left panel of Fig. 11 in Ref. [31]), leading to an *underestimate* of R_0 by roughly the same factor in order to keep the total number of observed events fixed.

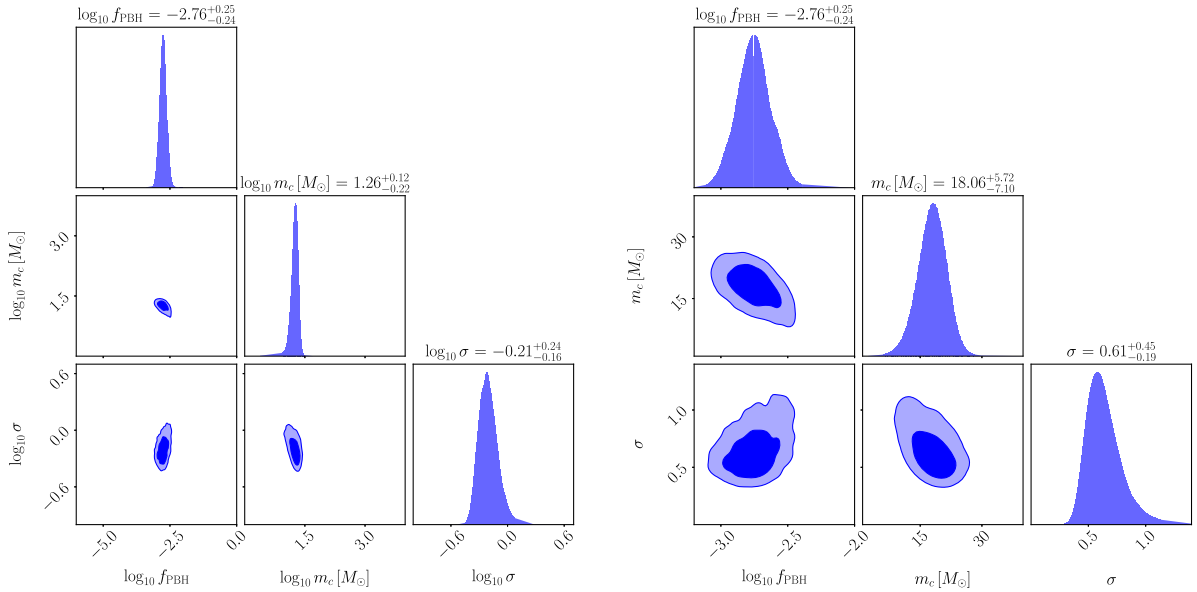

FIG. 8. *Left, off-diagonal panels*: Two-dimensional 68% and 95% marginal posterior quantiles for the parameters of the log-normal PBH model without the three-body suppression factor, given the GWTC-1 data. The plot boundaries correspond to the extent of the (uniform) priors on the parameters shown. *Left, diagonal panels*: One-dimensional marginal posterior densities for the parameters. Above each panel are the marginalized posterior median and 95% posterior quantiles for each parameter. *Right*: Same as left panel for the parameters $(\log_{10} f_{\rm PBH}, m_c, \sigma)$.

We close this section by noting that the weak constraints on the model B parameter β_q reflect the weak constraints on mass ratios in the GWTC-1 catalog (see Fig. 1). Recently, the LIGO-Virgo Collaboration reported detections of BBH mergers with significantly asymmetric masses having $q \approx 0.3$ [44] and $q \approx 0.1$ [45]. Consequently, the constraints on β_q tighten significantly when these sources are included. We only make use of the sources detected in the O1 and O2 observing runs in this work, but discuss the implications of reported O3 detections in Sec. VII. In a future work, we intend to repeat the analysis of this work with the $\mathcal{O}(100)$ detections expected in the final O3 catalog.

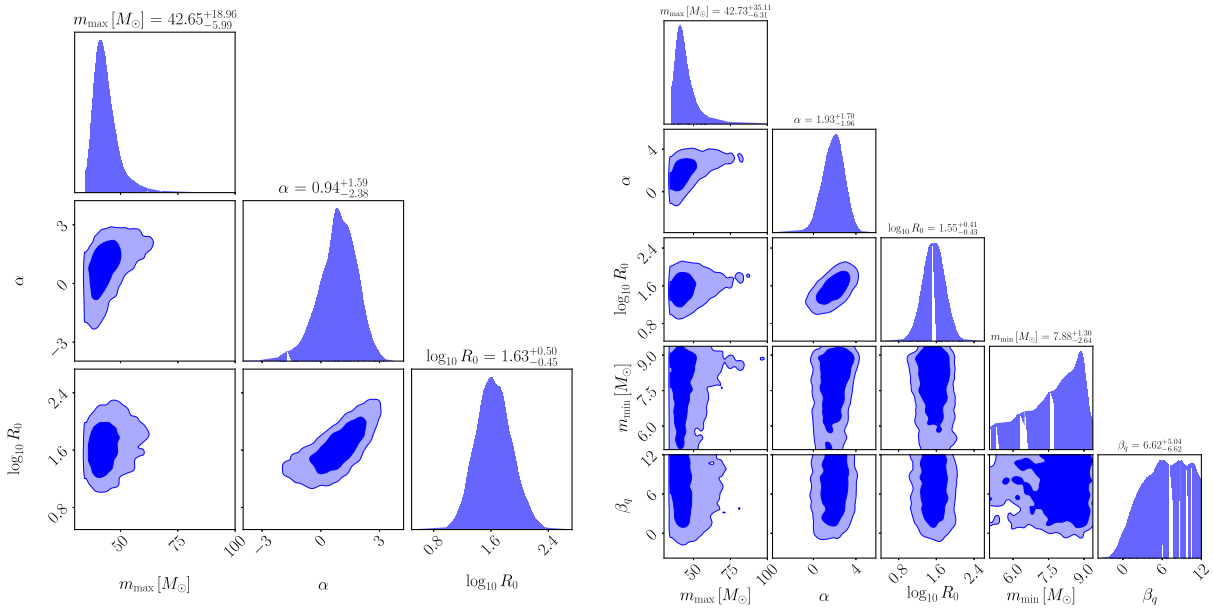

FIG. 9. Posteriors of the parameters of LIGO model A (left panel) and model B (right panel) given the GWTC-1 data. The meaning of the contours and quoted error significance are the same as in Fig. 7.

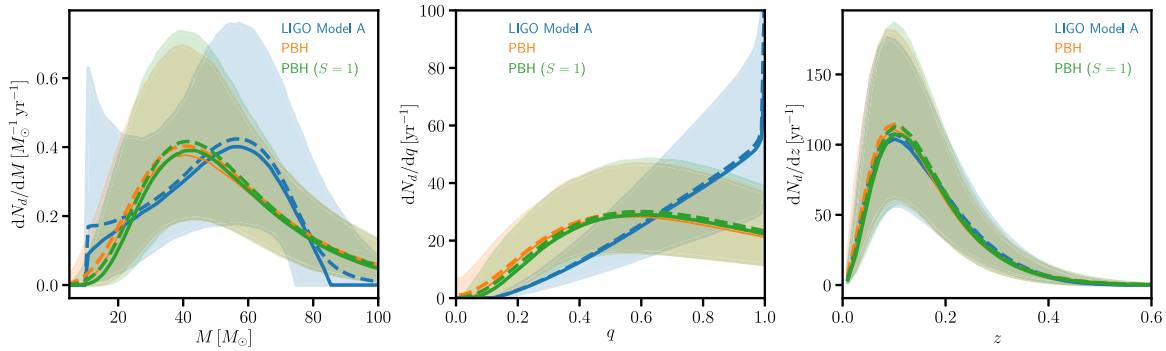

FIG. 10. Differential detector-frame merger rates with respect to total mass (left panel), mass ratio (middle panel), and redshift (right panel) for LIGO model A (blue), the log-normal PBH model (orange), and the log-normal PBH model with suppression factor set to unity (green). In each case, we plot the median and 90% quantiles over the posterior samples for each model given the GWTC-1 data (solid lines and shaded bands) and the (weighted) mean over the samples (dashed lines).

B. Evidences, goodness-of-fit tests, and model consistency

Having presented constraints on the parameters of the two PBH models and the two empirical LIGO models, we now examine the quality of model fits and compare the models using the Bayesian evidence.

1. Posterior merger rate distributions

We first examine the preferred distributions of source parameters in each model by computing the allowed values of the differential detector-frame merger rate, plotted in Fig. 10. These figures show the derivative of $N_d \equiv \beta/T_{\mathrm{obs}}$ with respect to total mass M, mass ratio q, and redshift z for each model averaged over the posterior distributions of the population hyperparameters. We show results for the PBH model, the PBH model with $S = 1$, and model A. The area under each curve in Fig. 10 is fixed at roughly $10/T_{\mathrm{obs}}$, since most of the posterior mass lies in a region where $\beta \approx 10$, matching the ten observed sources.

The preferred merger rates are similar among the three models, with the suppression factor making very little difference to the results. This is also evident from the parameter posteriors of the two models which have most of their mass in a similar region of parameter space, the degeneracy tail in the suppression factor model having little influence on the preferred differential merger rates. The redshift dependence in all three cases is determined from that of the selection probability p_{det} and the comoving volume element, cf. Fig. 3.

The distributions in total mass for the PBH models inherit the log-normal shape of the mass function $\psi(m)$, with a peak at roughly 40 $M_\odot$ and a long tail to high masses. Model A in contrast has both a minimum and maximum cutoff in mass. Note that the median model A merger rate is monotonically decreasing in its heavier mass ($\alpha \approx 0.4$), but this is counteracted by the detection probability p_{det} which increases with mass (see Fig. 2). These combined effects produce a peak in the merger rate around $M \approx 60\ M_\odot$ and a smaller peak around m_{min} where the differential merger rate formally diverges (visible in the 90% confidence region). In the case of the PBH models, the exponential falloff at high masses dominates over p_{det} giving a single peak.

The most pronounced differences occur in the dependence on mass ratio q (middle panel of Fig. 10). As pointed out in Ref. [21], the broad PBH mass function allows for mass ratios significantly different from unity. In contrast, and as is evident from comparing Figs. 6 and 4, model A merger rates typically have more symmetric masses. Despite the visible difference in the q distribution in Fig. 10, the errors on q in the GWTC-1 catalog are typically large, and the mass ratio has little discriminatory power between the PBH and LIGO models.[8]

2. Bayesian evidence ratios

In Table II, we present the Bayesian evidence for each model relative to that of model B. These quantities indicate the posterior preference for each model after marginalizing over each of its parameters.

For the PBH model, we find a (natural log) evidence ratio of -7.35 ± 0.23 compared to model B (errors here are approximately standard deviations). On the Jeffreys scale (e.g., Ref. [79]), this corresponds to "decisive" evidence in favor of model B compared with the PBH model. For the PBH model with $S = 1$, the log-Bayes factor is -8.25 ± 0.23; i.e., this model is even more disfavored compared with model B. The evidence ratio between the PBH models is 0.90 ± 0.23; i.e., the data do not show evidence for a suppression factor given our choice of priors.

For model A, we find a log-evidence ratio of -1.62 ± 0.22, i.e., positive or substantial but not strong evidence in favor of model B. This is consistent with the result reported

[8]We note that a source with $q \approx 0.3$ or $q \approx 0.1$, as recently detected in the O3 run, is predicted to be significantly more likely in the PBH models than under model A.

in Ref. [31] of -1.42. We note that Table 3 of Ref. [31] implies that the LIGO model C is slightly (but not significantly) preferred over model B. We find model A is strongly preferred over the PBH models with log evidences of -5.73 ± 0.23 and -6.63 ± 0.23 for the cases with and without the suppression factor, respectively.

Taken at face value, these evidences suggest that both PBH models are strongly disfavored compared with the simple empirical models A and B. However, it is well known that Bayesian evidences can be strongly influenced by the choice of priors, so it is beneficial to delve a bit deeper into the evidence ratios. We can make progress by employing the *Laplace approximation* for the evidence, discussed in Ref. [80]. This assumes that the posterior is approximately Gaussian around its peak (which occurs at the point $\boldsymbol{\theta}_{\rm BF}$), such that the integral over parameters can be approximated (for a uniform prior) as

$$p(\mathbf{d}|M) \approx p(\mathbf{d}|\boldsymbol{\theta}_{\rm BF}, M) \times \frac{\sqrt{\det(2\pi\mathbf{C})}}{\mathrm{Vol}_{\pi(\boldsymbol{\theta})}}, \tag{23}$$

where $\mathbf{C}$ is the covariance matrix of the posterior and $\mathrm{Vol}_{\pi(\boldsymbol{\theta})}$ is the prior volume (i.e., the volume of the cube defining our uniform priors). The first term on the right-hand side of Eq. (23) is the likelihood value at of the best-fitting model, a quantifier of model fit quality well known from classical statistics. The second term is the "Occam factor" expressing the ratio of the posterior volume to the prior volume. The Occam factor quantifies the degree to which the region of acceptable parameter values shrinks upon arrival of the data and penalizes models for which this shrinkage is large, i.e., models which require finely tuned parameter values among those which were allowed *a priori*.

Since the posterior of both model B and the PBH model with suppression are significantly non-Gaussian, the Laplace approximation is expected to be only a coarse model for the evidence. In Table II, we give the differences between the Laplace-approximated evidence $Z_{\rm Lap}$ and the nested sampling estimate Z_{NS}. We find that $Z_{\rm Lap}$ provides a remarkably good approximation to the true evidences, with log ratios ranging from roughly 0.26 for the $S = 1$ model (which has the most Gaussian posterior) to 1.60 for the full PBH model (which has a strongly non-Gaussian posterior). In all cases, the discrepancy between $Z_{\rm Lap}$ and Z_{NS} is significantly smaller than the difference from the evidence for model B.

With the Laplace-approximated evidences, we can start to understand why some models are favored over others. In Table II, we show the ratios of the terms in Eq. (23) with those of model B. The ratio of first terms is just L^*/L_B^*, the likelihood ratio of the best-fit model compared with that of model B. In the case of the PBH models, we find that this term dominates the evidence ratio. The Occam factor is similar among the PBH model with suppression factor and model B but is more penalizing for the $S = 1$ PBH model since the shrinkage in prior volume is much greater, as evident from Fig. 8.

The log-Bayes factors depend on the prior volume, via the Occam factor, as $\ln Z \sim -\ln \mathrm{Vol}_{\pi(\boldsymbol{\theta})}$. The evidence ratios of the PBH models compared with the LIGO models are thus sensitive to the prior range on $f_{\rm PBH}$ and the mass function parameters. If, for example, we reduced the prior lower limit on $f_{\rm PBH}$ from 10^{-6} to 10^{-16}, we would reduce the log-Bayes factor compared to model B by roughly 1.[9] This prior range could be easily exceeded if instead we placed a uniform-in-log prior on the primordial power spectrum amplitude, to which $f_{\rm PBH}$ is exponentially sensitive, a point we discuss further in Sec. VII. Similarly, increasing the ranges of m_c and σ would also increase the evidence against the PBH models compared with the LIGO models.

Note that there is no freedom to *reduce* the prior range of the PBH model parameters without being overly informative; i.e., we cannot attempt to boost the evidence of PBH models by making the Occam factor less penalizing, unless some other prior information or physical insight demands it. Could we instead try to penalize the LIGO models to restore the prior ambivalence between models? In the case of model A, we would need to increase the prior volume by a factor of roughly 300 to give an evidence ratio of unity with the PBH model. This could be achieved by expanding each side of the prior cube by a factor of roughly 6.7, i.e., with priors $m_{\rm max}[M_\odot] = [30, 500]$, $\log_{10} R_0 = [-12, 14]$, and $\alpha = [-50, 58]$. *A priori*, this seems an extreme prior range which is likely to be unphysical. Rather than change the limits of the prior, we could also change its density such that the prior volume contained more prior mass [note this would require a modification to Eq. (23)]. In the absence of a more fundamental astrophysical theory, there is no obviously preferred choice of parameter combination on which to impose a uniform prior. There is thus no well-motivated way to make the data favor the PBH model over the LIGO models by simply changing parameter priors.

We note that the prior volumes cancel in the evidence ratio of the PBH model with and without a suppression factor. We can therefore make the robust statement that the GWTC-1 data are not sensitive to the suppression factor, and inference of PBH models may proceed with $S = 1$ with negligible loss of accuracy.

3. Posterior predictive distributions

Having seen that the likelihood ratio is primarily responsible for the evidence against a log-normal PBH model, and having argued that changing the prior on parameters to restore model parity is challenging, we

[9] In reality, our posterior limits on $f_{\rm PBH}$ are upper limits due to our assumption that every source is a PBH-PBH merger, so the Occam factor is probably not as penalizing as this example suggests.

now investigate the cause of the likelihood differences in detail. Equation (23) tells us that it is sufficient to consider only the likelihood at the best-fit model, but the same conclusions can be reached by considering the likelihood averaged over the model space allowed by the data—the *posterior predictive distribution* (PPD), defined as

$$p(D|\mathbf{d}, N_{\rm obs}, M) = \int \mathrm{d}\boldsymbol{\theta}\, p(D|\boldsymbol{\theta}, M)\, p(\boldsymbol{\theta}|\mathbf{d}, N_{\rm obs}, M), \quad (24)$$

where D is *unseen* data. Equation (24) is similar to the Bayesian evidence, except it is now an integral of the likelihood of new data over model parameters allowed by the old data $\mathbf{d}$ and $N_{\rm obs}$.

The PPD is a useful quantity to compute since it can be used to approximate the part of the evidence ratio coming from the likelihood ratio in the Laplace approximation (23). It can also be used to assess the absolute quality of the model fit in a more "Bayesian" way than a classical χ^2 test [81].

We approximate the integral in Eq. (24) with an average over posterior samples from our nested sampling runs. The challenge in implementing the PPD is finding an approximation for the likelihood of unseen data $p(D|\boldsymbol{\theta}, M)$, as so far we have only needed the likelihood of the GWTC-1 data as a function of parameters, which we extracted indirectly via the GWTC-1 posterior samples. In Appendix C, we present a detailed derivation of this approximate likelihood. This results in a PPD for the redshifted chirp mass given by

$$p(\{\widehat{\mathcal{M}}_z\}|\{\mathbf{d}\}) = \prod_{i=1}^{N_{\rm obs}} \int \mathrm{d}\mathcal{M}_z\, p(\widehat{\mathcal{M}}_z^i|\mathcal{M}_z)\, p(\mathcal{M}_z|\{\mathbf{d}\}), \quad (25)$$

where

$$p(\mathcal{M}_z|\{\mathbf{d}\}) = \int \mathrm{d}\boldsymbol{\theta}\, p(\mathcal{M}_z|\boldsymbol{\theta})\, p(\boldsymbol{\theta}|\{\mathbf{d}\}). \quad (26)$$

In these expressions, $\{\widehat{\mathcal{M}}_z\}$ denotes the set of "measured" chirp masses, with members of the set denoted by $\widehat{\mathcal{M}}_z^i$.

The PPD in Eq. (26) can be convolved with the individual constraints on chirp mass from each source to give a likelihood for new unseen chirp masses averaged over population parameters allowed by existing data. When evaluated at the actual chirp mass values in the GWTC-1 catalog, this gives the likelihood function marginalized over the absolute merger rate and the population parameters—this is approximately equivalent to the likelihood evaluated at the best-fitting population model, which is the key quantity in determining whether the Bayesian evidence favors PBH over the LIGO models.

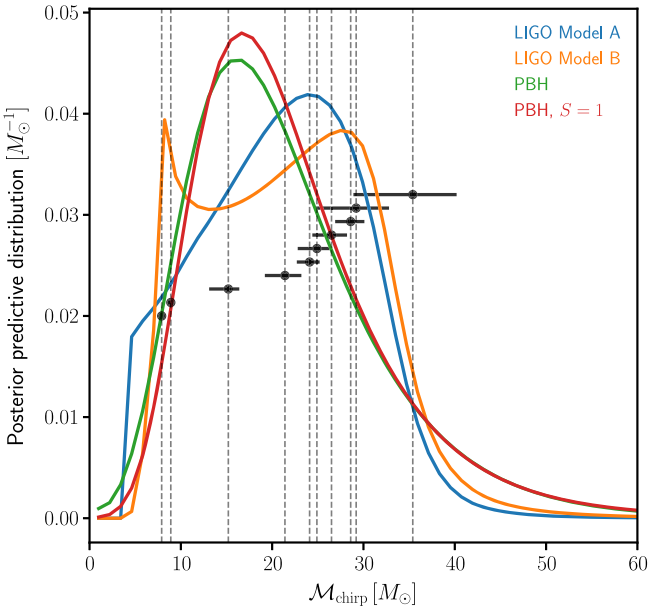

FIG. 11. PPD of the source-frame chirp mass, given the data, for LIGO model A (blue), LIGO model B (orange), the log-normal PBH model (green), and the log-normal PBH model with suppression factor set to unity (red). Note that there is a thin spike with width approximately $10^{-3}\ M_\odot$ at $\mathcal{M}_{\rm chirp} \approx 4.35\ M_\odot$ where the PPD diverges in the case of model A, not visible on this plot due to the resolution. This spike corresponds to $m_1 \approx m_2 \approx m_{\rm min}$ and gives negligible contribution to the integrated PPD. We also show the median (black vertical lines and points) and 90% credible intervals (black horizontal bars) of the source-frame chirp masses of the GWTC-1 BBH sources [16] (with an arbitrary vertical offset for visual clarity).

In Fig. 11, we show the PPD on source-frame chirp mass[10] given by Eq. (26), along with the central values and 90% confidence intervals for the sources in the GWTC-1 catalog—the best-fit likelihood for each model is approximately the PPD plotted in the figure convolved with each of the source posteriors and then evaluated at their central values. Ignoring the uncertainties in observed chirp mass, this simply amounts to recording the height of the PPD curves where they intersect each of the observed values. A model with a peak in its PPD located in the vicinity of a large number of measured chirp masses will have a higher likelihood than a model peaking away from where the observations are. This directly translates into a higher Bayesian evidence via the Laplace approximation (23). Put even more coarsely, the likelihood ratio is effectively comparing the coherence of the empirical histogram of chirp masses with the predicted distribution for each model.

[10]The detector-frame chirp mass PPD is very similar due to the low redshifts of the sources, but is significantly more computationally expensive to generate due to the suppression factor of the PBH model.

The key features of the curves in Fig. 11 are similar to those in the left panel of Fig. 10 where we plotted the equivalent distribution for total mass. The suppression factor makes little difference to the PBH model, which demonstrates a log-normal shape in the chirp mass distribution inherited from the mass function $\psi(m)$. Model A and model B both have sharp peaks around $m_1 \approx m_2 \approx m_{\rm min}$, with the peak in model A narrower than the resolution of the plot, having width approximately $10^{-3}\ M_\odot$. These peaks are due to a formal divergence in the merger rate caused by the $C(m_1)$ term in Eq. (12), i.e., the requirement that the marginal distribution in the heavier mass be a power law. The secondary peak around $\mathcal{M}_{\rm chirp} \approx 30\ M_\odot$ is due to the detection probability. No such peak is seen in the PBH model due to its more extreme falloff with increasing chirp mass.

The combined effect of the merger rate and the detection probability is that the PPD of model B is able to peak sharply at the location of the two well-measured light binaries with $\mathcal{M}_{\rm chirp} \lesssim 10\ M_\odot$, predicting fewer sources in the range $10\ M_\odot$–$20\ M_\odot$ in agreement with observations, before peaking again in the region $20\ M_\odot$–$30\ M_\odot$ just where the majority of the measurements are. Model A can also do this to a lesser extent, but is disfavored compared with model B because it gives less likelihood to the two light sources. This is simply a reflection of the fact that model B has the freedom to fit the minimum component mass $m_{\rm min}$. Since there will always be a sharp peak in the chirp mass distribution at the minimum chirp mass, it is always advantageous for a model to place $m_{\rm min}$ as close to the actual minimum mass as possible. The penalty incurred from the Occam factor in this fine-tuning process is substantially outweighed by the increase in likelihood. Model A, in contrast, has a fixed $m_{\rm min}$.

Turning now to the PBH models in Fig. 11, it is clear that a log-normal distribution will struggle to fit the observed distribution of chirp masses compared with models A and B. The mass function parameters m_c and σ are fit to ensure the log-normal peaks in the correct mass range and has a width encompassing the observed range of values, but the detailed shape is a poor fit to the data even with only ten points. The LIGO models are able to fit the key features of the empirical distribution, namely the high density of chirp masses in the $20\ M_\odot$–$30\ M_\odot$ region; the relative dearth in sources between $10\ M_\odot$ and $20\ M_\odot$; and, in the case of model B, the two light sources with $\mathcal{M}_{\rm chirp} \lesssim 10\ M_\odot$.

The actual likelihoods at the observed data points are given approximately by the values of the curves where they intersect the vertical dashed lines in Fig. 11. While the PBH models intersect the sources at approximately $15\ M_\odot$ and $35\ M_\odot$ at higher values than both LIGO models, they both fail to capture the cluster of sources in the $20\ M_\odot$–$30\ M_\odot$ region. PBH models having a peak in this region are not as favored as those having a peak around approximately $15\ M_\odot$, since they typically overpredict sources at heavier mass compared with lighter mass. The likelihood ratio (and hence the evidence ratio under the Laplace approximation) penalizes the PBH models precisely for this reason. If new data populated Fig. 11 with many binaries having chirp masses greater than $35\ M_\odot$, we would expect the log-normal PBH models to perform relatively better since they naturally predict a long positive tail in the distribution. The absence of sources above $35\ M_\odot$, readily detectable in O1 and O2, penalizes PBH models which predict they should be there.

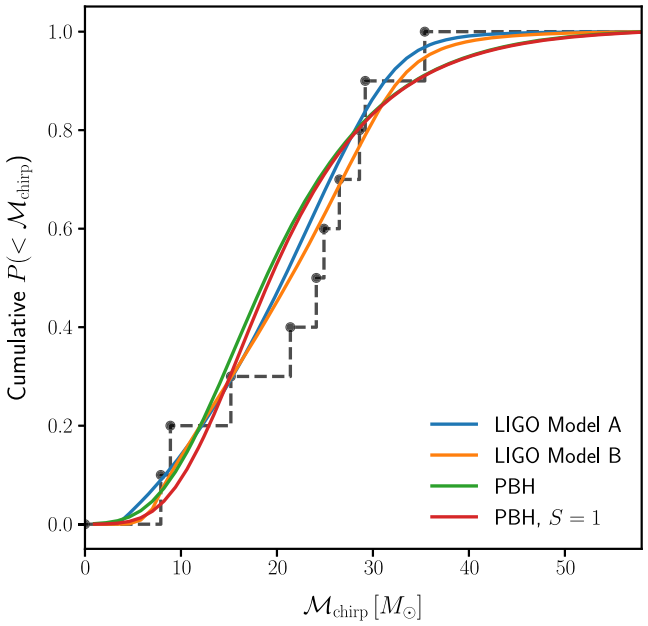

FIG. 12. Posterior CDF of source-frame chirp mass, given the data, along with the empirical CDF of the sources, for LIGO model A (blue), LIGO model B (orange), the log-normal PBH model (green), and the log-normal PBH model with suppression factor set to unity (red).

An alternative way of looking at the differences between the PPD and the measured chirp mass is in the cumulative version of the PPD [posterior cumulative distribution function (CDF)] found by integrating from zero mass up to some specified value. The CDF for the LIGO models in terms of heavier mass was also studied in the recent Ref. [82]. In Fig. 12, we show this quantity for the PBH models along with models A and B.

The CDF is constrained to lie between zero and unity, and the model curves show significant overlap when plotted in this way. All models overpredict the number of mergers with chirp mass less than or approximately equal to $20\ M_\odot$, which can be seen as the relative dearth of sources in the $10 - 20\ M_\odot$ region of Fig. 11, with model B clearly performing the best. Note that this behavior is not evident in the CDF of the heavier mass plotted in Ref. [82]. We have argued that the chirp mass is the more appropriate parameter to use since it is less correlated with other parameters, is the best constrained source parameter to

leading order, and has a PPD most directly related to the likelihood and the evidence.

The CDF permits the use of a Kolmogorov-Smirnov (KS) test using the empirical CDF (shown as the gray lines and points in Fig. 12). However, this test looks only at the maximum deviation of the predictions from the data (occurring around 20 $M_\odot$), and the p-values from a one-sample KS test are all between 0.3 and 0.8, indicating that all models are acceptable fits to the data. This is not as powerful a test as the likelihood ratio or Bayesian evidence, however, which uses the detailed shape of the chirp mass distribution to assess its ability to fit the data.

To summarize the results of this section, we have seen that the log-normal PBH models are significantly disfavored compared with the empirical LIGO models, quantified by the Bayesian evidence ratio. We have shown that the evidence ratios between each model can be well approximated by the product of a likelihood ratio and an Occam factor. The Occam factor is sensitive to the prior volumes, and the evidence ratio can be made to restore the prior ambivalence toward all models by broadening the priors on the LIGO models, but extreme values must be imposed to achieve this. The likelihood ratios are the dominant source of evidence against the PBH models, and we have shown how this can be reduced to the ability of models to predict the empirical distribution of chirp masses in the GWTC-1 catalog. One of the main results of this work, Fig. 11, demonstrates that the log-normal mass function struggles to match the detailed distribution of observed chirp mass, predicting positive skewness when the data appear to prefer negative skewness. In contrast, model A can predict negatively skewed chirp mass distributions and has consequently higher likelihood. Model B can additionally predict the two low-mass events and the relative dearth of objects at intermediate masses and is favored over model A. We caution that the LIGO models have some features which lack strong physical motivation and hence we do not advocate that the log-normal PBH model should be abandoned in favor of these models. We have instead shown why models in which every source is a PBH-PBH merger struggle in comparison, and identified the features of the data that need to be explained if the log-normal mass function is to become favorable. LIGO model C is both more physical and a better fit to the data than models A and B, and therefore by extension is significantly preferred (in terms of the Bayes factor) over the PBH models. It thus seems almost certain that successful PBH models will either need to abandon the prediction that every merger detected by LIGO and Virgo is a PBH merger or introduce a physical mechanism that significantly modifies the primordial log-normal mass function.

VI. EXTENSIONS TO THE LOG-NORMAL PBH MASS FUNCTION

We have seen that PBH models with a log-normal mass function do not provide as good a fit to the LIGO data as simple empirical models. Since this family of mass functions is highly constrained, having only two free parameters in addition to an overall normalization, we now study simple extensions to $\psi(m)$ to investigate whether a better fit might be achieved with minimal modification.

We set the suppression factor equal to unity for all extended models considered in this section. The results of Sec. V showed that the suppression factor has only a modest influence on the preferred models while greatly increasing the run time of the likelihood calculation, so for simplicity, we set $S = 1$ henceforth.

A. Log-normal with a high mass cutoff

We argued in Sec. V that one of the reasons a log-normal struggles to fit the observations is its long positive tail to high chirp masses, not seen in the data. By comparison, both the LIGO models we consider have explicit cutoffs at high component masses. We therefore now consider a new mass function $\psi_{\rm cut}$, defined by

$$\psi_{\rm cut}(m) \propto \begin{cases} \frac{1}{m\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\ln^2(m/mc)}{2\sigma^2}\right] & m \leq m_{\rm max} \\ 0 & m > m_{\rm max} \end{cases} \quad (27)$$

where the normalization is chosen such that $\psi_{\rm cut}$ integrates to unity. We note that such an upper mass cutoff is difficult to construct in PBH formation models.

For simplicity, and in order to approximately maximize the evidence for the PBH models without adding a penalty for adding a new free parameter, we fix $m_{\rm max} = 50\ M_\odot$, such that the maximum source-frame chirp mass is $\mathcal{M}_{\rm chirp} \approx 44\ M_\odot$. Figure 11 shows that this lies just beyond the 90% upper chirp mass of the heaviest source, which implies that this choice of cutoff is not too restrictive.

We rerun the parameter inference and evidence calculation on the GWTC-1 catalog data using this model, with the suppression factor set to unity and the same priors on $f_{\rm PBH}$, m_c, and σ as the log-normal model (i.e., uniform in the log of each parameter). The resulting parameter posteriors are shown in Fig. 13.

The constraints on $f_{\rm PBH}$ in this model are almost the same as the no-cutoff case, with values preferred which give $\beta \approx 10$ events. The main difference comes in the preferred values of m_c and σ, with Fig. 13 revealing a long degeneracy tail stretching to high values of m_c and σ. This is caused by the high-mass cutoff permitting values of $m_c \gg m_{\rm max}$ if σ is sufficiently large that there is still a high likelihood of mergers happening in the observed mass range. When σ is small enough, m_c corresponds to this observed mass range and is constrained to similar values as in the absence of a cutoff.

The log evidence of the cutoff model compared to model B is given in Table II and is -4.01 ± 0.21. This model is thus strongly preferred over the models without a cutoff. Model A and model B are both still strongly preferred over

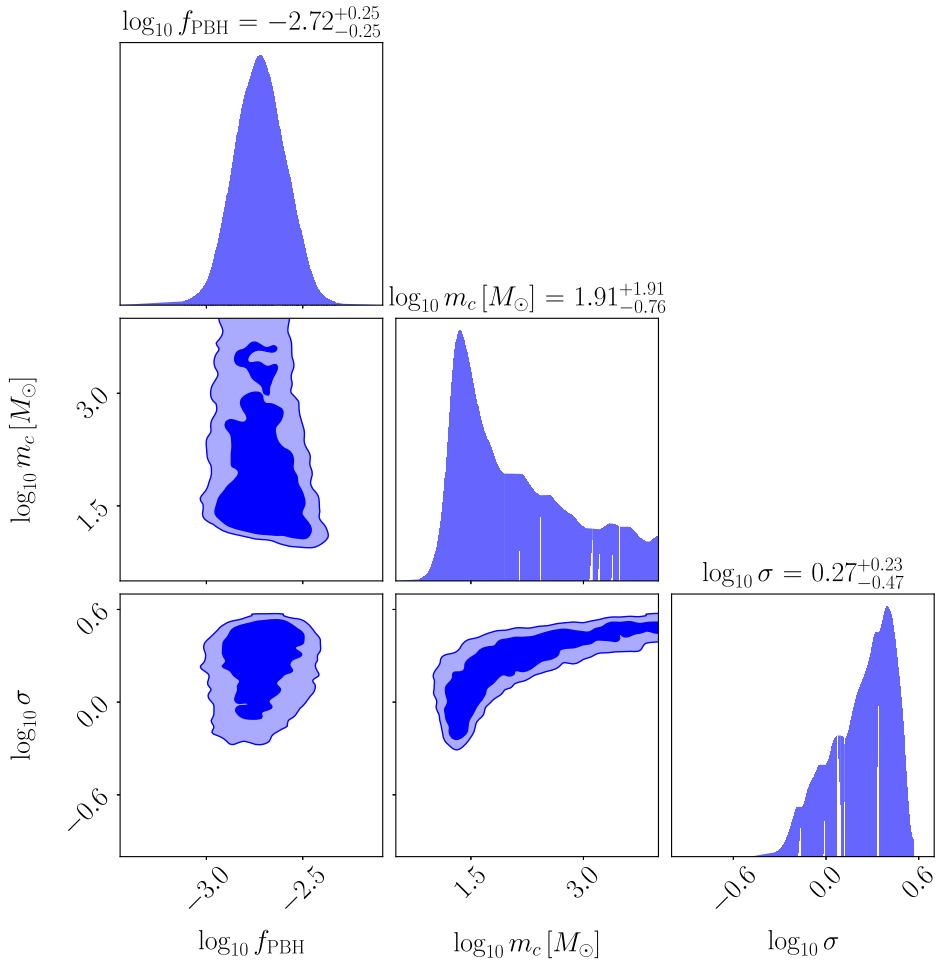

FIG. 13. Posteriors of the log-normal model without the suppression factor and imposing a maximum cutoff mass (source frame) of $m_{\rm max} = 50\ M_\odot$. The meaning of the contours and quoted error significance are the same as in Fig. 7.

this cutoff model. The Laplace-approximated evidence ratio differs from the nested sampling estimate by only 0.54 ± 0.13, but due to the highly non-Gaussian posterior, we do not expect this to be accurate. Nevertheless, this approximation implies that the evidence ratio is dominated by the likelihood ratio against model B, with the Occam factor now less penalizing due to the larger posterior volume permitted by the data (compare Fig. 13 with Fig. 8). We note that had we allowed $m_{\rm max}$ to vary and be constrained by the data this conclusion might change, due to the large prior volume that could be assigned to $m_{\rm max}$. We note, however, that with fixed $m_{\rm max}$ the evidence ratios compared to the no-cutoff models are independent of the priors, which are the same among these PBH models.

In Fig. 14, we show the PPD of the source-frame chirp mass for this model (purple curve) along with that of LIGO models A and B. With the preferred values of m_c skewed to values much greater than $m_{\rm max}$, the shape of the posterior-averaged likelihood now looks quite different than the log-normal case. In particular, the large positive skewness has been suppressed by the cutoff, and the distribution is approximately symmetric about its peak at roughly $23\ M_\odot$. As in the log-normal case, the preferred values of the parameters m_c and σ are such that the overall distribution has roughly the correct absolute mass scale and a width incorporating the observed chirp masses. It is clear from Fig. 14 that the broader distribution allowed by the cutoff mass function is a better fit to the data and accounts for the increased evidence for this model. Despite this, there is thus no additional freedom to fit the detailed distribution of the data beyond m_c and σ, and model B still provides a better fit. Model A is also preferred to the cutoff log-normal model, due in part to its low-mass cutoff which allows the likelihood of the two well-measured low-mass sources to be higher.

B. Skew-log-normal

While a log-normal mass function for PBHs may be shown to be an excellent approximation to a wide range of peaklike features in the primordial power spectrum, for very narrow peaks, a negatively skewed log-normal is

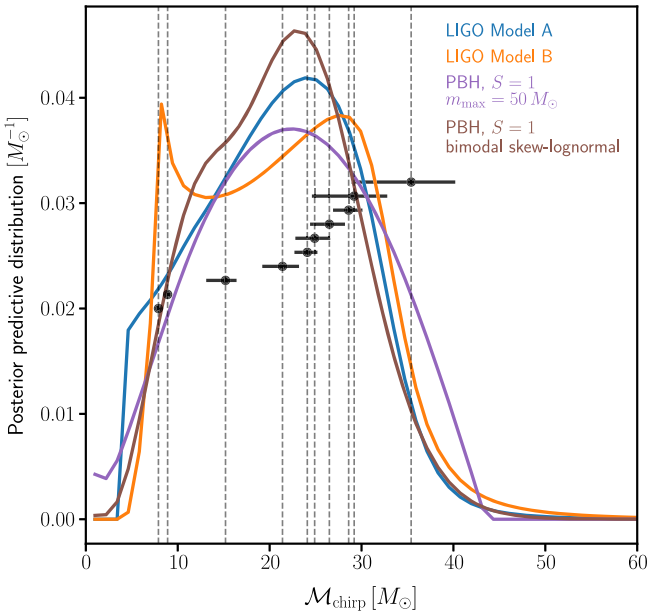

FIG. 14. PPD for two extensions to the log-normal PBH mass function: the $S = 1$ log-normal model with a cutoff $m_{\rm max} = 50\,M_\odot$ (purple) and the bimodal skew-log-normal model with $S = 1$ (brown). We also show LIGO model A (blue) and model B (orange), which are the same as in Fig. 11.

a better approximation [62]. We thus consider a skewed log-normal mass function given by

$$\psi_{\rm skew}(m) = [1 + {\rm erf}(\alpha\Delta)]\psi(m), \tag{28}$$

where $\psi(m)$ is a log-normal mass function and $\Delta \equiv \ln(m/m_c)/(\sqrt{2}\sigma)$ is the logarithmic mass deviation. The skewness is parametrized by a parameter α controlling the argument to the error function ${\rm erf}(x)$. As shown in Ref. [62], a delta function in the power spectrum corresponds to $\alpha \approx -2.6$ and $\sigma \approx 0.56$, which we impose here. This leaves $f_{\rm PBH}$ and m_c as the only free parameters of this model.

Running the nested sampling inference with the skew-log-normal model, we find the best-fitting values of the parameters are $f_{\rm PBH} = 1.4 \times 10^{-3}$ and $m_c = 34.0\,M_\odot$, comparable with the results found for the log-normal model. The (log) likelihood ratio at the best-fit compared to model B is -8.3, i.e., very similar to the full log-normal ($S = 1$) case. This is due to the relatively weak skewness of the model and the fact that $\sigma = 0.56$ is actually quite close to the value preferred by the data.

The evidence for the model relative to model B is -7.80 ± 0.23, i.e., slightly preferred compared with the nonskewed PBH model with $S = 1$ but still strongly disfavored compared to both model A and model B. The model provides a fit to the data comparable with the unskewed log-normal mass function, i.e., not competitive with the LIGO empirical models. A Laplace approximation to the evidence is very accurate and shows that the increased evidence for the model results from a less penalizing Occam factor due to its reduced dimensionality compared with the nonskewed model. However, this change is not enough to overcome the big difference in likelihood ratio which gives rise to strong evidence against the model compared with the LIGO models.

C. Bimodal skew-log-normal

The distribution of measured chirp masses in the GWTC-1 catalog has a cluster of sources with $\mathcal{M}_{\rm chirp} \approx 30\,M_\odot$, a relative dearth of sources between $10\,M_\odot$ and $20\,M_\odot$, and two well-measured light sources with $\mathcal{M}_{\rm chirp} < 10\,M_\odot$. Motivated by this, we consider a mixture of two skew-log-normal mass functions for PBHs given by

$$\psi_{\rm skew,bi}(m; m_{c,1}, m_{c,2}) = \lambda\psi_{\rm skew}(m; m_{c,1}) + (1-\lambda)\psi_{\rm skew}(m; m_{c,2}), \tag{29}$$

where $\psi_{\rm skew}$ is the skew-log-normal distribution introduced in Sec. VI B; i.e., each component has fixed skewness parameter $\alpha = -2.6$ and scale parameter $\sigma = 0.56$. Such a mass function could arise from two distinct narrow peaks in the primordial power spectrum whose amplitudes must be tuned if a comparable number of PBHs are to be generated by each peak [22,83]. However, we note that very close peaks will not produce the distribution (29) in detail due to the two peaks "smearing" into each other.

We choose log-uniform priors in location parameters $m_{c,1}$ and $m_{c,2}$ with limits given in Table I. To avoid redundant likelihood calculations implied by the symmetry of Eq. (29), we impose a uniform prior on λ in the range $[0, 0.5]$, such that $m_{c,2}$ is defined as the location parameter of the dominant component in the mixture.

In Fig. 15, we show the posterior constraints on the parameters of the skew-bimodal model. As for the other model extensions studied in this section, the preferred values of $f_{\rm PBH}$ are such that the total number of events is roughly 10, the median value being $f_{\rm PBH} = 0.002$ in this model. The distribution in the $[m_{c,1}, m_{c,2}]$ plane is bimodal, with a dominant peak at $m_{c,1} \approx 10\,M_\odot$ and $m_{c,2} \approx 35\,M_\odot$, which matches our expectation given the observed chirp masses; for $q = 1$, this corresponds to a dominant component in the mass function at $\mathcal{M}_{\rm chirp} \approx 30\,M_\odot$ and a subdominant component at $\mathcal{M}_{\rm chirp} \approx 9\,M_\odot$. The subdominant peak in the mass posterior corresponds approximately to swapping which of these mass function peaks is dominant, preserving their location.

The mixture parameter λ is poorly constrained, with values around $\lambda \approx 0.4$ typically preferred. Bimodality in the posterior appears when $\lambda \gtrsim 0.35$, which is to be expected since the data are not constraining enough to distinguish which mass function peak is dominant when the difference

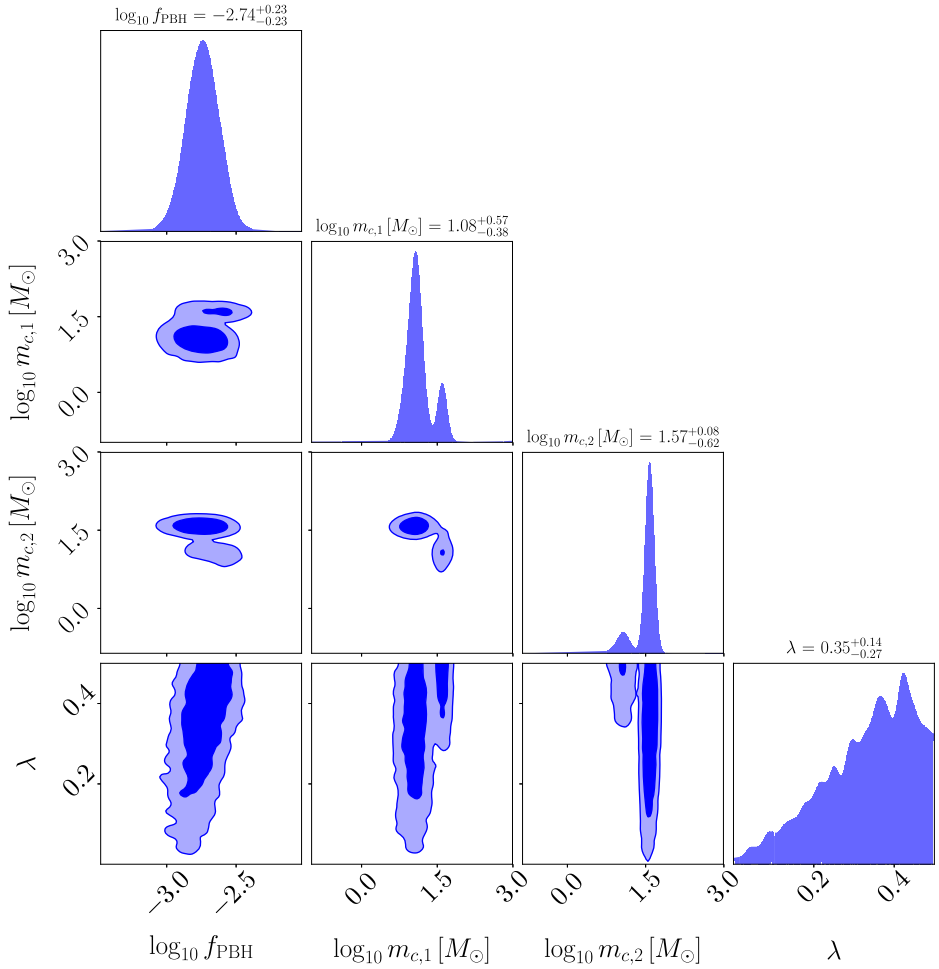

FIG. 15. Posteriors of the bimodal skew-log-normal model without the suppression factor, fixing the shape parameter of each component to $\sigma = 0.56$ and the skewness parameter to $\alpha = -2.6$, roughly corresponding to delta functions in the power spectrum. The meaning of the contours and quoted error significance are the same as in Fig. 7.

is sufficiently small, i.e., when λ is sufficiently close to the point of symmetry at $\lambda = 0.5$.

The evidence for this model is reported in Table II and is -5.79 ± 0.24 compared to model B. There is thus "substantial" evidence (on the Jeffreys scale) for this model compared with both the PBH log-normal models considered previously, although the Bayesian evidence in favor of models A and B is still comparatively strong. The likelihood ratio to model B is much more favorable for the skew-bimodal model compared with the other PBH models and is only marginally smaller than model A. The Occam factor is comparatively more penalizing than both the LIGO models, although the highly non-Gaussian parameter posterior makes the Laplace approximation a poor estimate of the true evidence.

In Fig. 14, we show the PPD for the source-frame chirp mass in this model (brown curve). The peak at small masses has been skewed to heavier masses by the detection probability, but the two components are still clearly distinguishable in this plot. The comparable likelihood ratio between the skew-bimodal mass function and model A (blue curve) is clear from Fig. 14, with the additional low-mass component matching model A's ability to assign high likelihood to the two well-measured light BBHs. The PPD suggests that the skew-bimodal model performs less well compared with model B due to its inability to predict a sharp peak at low chirp masses. This could potentially be remedied by allowing one or both of the variance parameters in the skew-log-normal components to vary, although this would come at the price of a more penalizing Occam factor (i.e., overfitting the data).

The skew-bimodal model is thus successful at matching the fit of the observed distribution of chirp masses provided by the LIGO models and is the most successful of the PBH models we consider. We note that the two preferred central values of the components are reasonably close, such that this model might not be expected to be an accurate approximation to two delta-function peaks in the primordial power spectrum. We also note that constructing a physical mechanism that could produce two such peaks of

comparable amplitude is not straightforward and the *a priori* motivation for this is weak.

It is perfectly possible of course that the apparent peak at low chirp masses is not a "real" feature of the population but a consequence of the low-number statistics. The Bayesian evidence accounts for this, but the likelihood ratio does not, so we caution against attempting to construct models to fit the detailed empirical distribution of chirp masses in general. Nonetheless, we have seen that Bayesian evidence favors the skew-bimodal model over a single log-normal component.

D. Late-time PBH capture model

The final merger rate model we consider is a "late-time capture" model in which PBH binaries form in the late Universe via two-body encounters. We adopt the model of Refs. [12,84] in which the differential merger rate is given by

$$\frac{\mathrm{d}R}{\mathrm{d}m_1\mathrm{d}m_2} \propto \frac{(m_1+m_2)^{10/7}}{(m_1 m_2)^{5/7}}\psi(m_1)\psi(m_2). \tag{30}$$

This model follows from the analytical calculations of Refs. [85,86] which model two compact objects on an initially parabolic or hyperbolic trajectory which become bound due to the radiation of gravitational waves, using an accurate quasi-Newtonian approximation. Reference [12] additionally assumes that the relative velocity of the two objects is independent of their masses, an assumption which may break down in detail due to mass segregation in halos. The prefactor of Eq. (30) is a free parameter in this model which can be high enough to give the observed merger rate even for low f_{PBH}, due to enhancements in the merger rate within dense halos. We assume that the mass functions in Eq. (30) are log-normals with parameters m_c and σ, having the same priors as in our baseline PBH model.

The best-fit values for m_c and σ were obtained with a numerical optimization routine and are 16.5 $M_\odot$ and 0.56, respectively, i.e., almost identical to the baseline $S=1$ PBH model. The best-fit mass function has a shape in the (m_1, m_2) plane almost identical to that of the baseline early-time formation model, differing only in the tails. This is due to the low value of σ preferred by the data, which keeps the mass function compact and suppresses the influences of mass-dependent prefactors multiplying the log-normal $\psi(m_1)\psi(m_2)$ term in the merger rate. Given the similarity with the early-time model, we chose not to run the full nested sampling inference on the late-time capture model.

This model thus provides a very similar fit to the early-time model, with the quality of the fit dominated by the log-normal shape of the mass function. This may be compared to the result that the suppression factor makes little difference to the best-fitting PBH model and its maximum likelihood and suggests that radical mass-dependent evolution of the initial PBH mass function is required to give a good fit to the data when the mass function is log-normal.

VII. CONCLUSIONS

In this work, we have confronted the latest PBH binary merger models with the catalog of gravitational wave merger events from the first two observing runs of LIGO-Virgo. We have adopted a Bayesian formalism throughout, which has allowed us to place posterior probabilities on the parameters of the PBH model given the data, accounting for the source parameter correlations and the interferometer selection function. Assuming all the observed black hole mergers are primordial and marginalizing over the mass function parameters, we find $f_{\mathrm{PBH}} = (5.0^{+67.4}_{-2.8}) \times 10^{-3}$ (median and 95% confidence), with a long positive tail allowed by the data due to a suppression in the merger rate from demanding the binary is not disrupted by other PBHs. Relaxing this requirement gives the smaller value $(1.7^{+1.4}_{-0.7}) \times 10^{-3}$. The preferred mass function parameters are such that the observed black hole mass scale and spread in masses is correctly predicted. A log-normal fit gives a central value of approximately 20 $M_\odot$ and a logarithmic width of order unity, consistent with previous results [14,26,34].

Going beyond parameter constraints, we have studied the quality of the model fits using several Bayesian tests. We computed the Bayesian evidence for the PBH models and popular astrophysically motivated models, finding in all cases that the astrophysical models were favored decisively. By making a Laplace approximation, we decomposed the evidence ratio into a likelihood ratio (well known from classical frequentist statistics) and an Occam factor quantifying the sensitivity to the parameter priors. This exercise showed that the evidence ratio was dominated by the likelihood ratio, i.e., by the relative goodness of fit of the best-fitting models in the PBH and astrophysical scenarios. We were able to show that this may be understood by comparing the predicted distributions of the chirp mass with the observed values and identified the posterior predictive distribution as a crucial descriptor of the relative quality of the model fits.

Using the posterior predictive distribution, we were able to show that PBH models struggle because they predict a chirp mass distribution with a close-to-log-normal shape, in marked contrast to the observations. The LIGO interferometers were sensitive enough in their first two observing runs to detect black holes with chirp masses well beyond the 40 $M_\odot$ upper limit of the sources which were actually detected. A log-normal distribution has a long positive tail to high masses, overpredicting the abundance of high-mass binaries. Likewise, the detailed distribution of the observed sources is not well predicted compared with the empirical astrophysical models. These empirical models are parametrized in terms of the heavier mass and the mass ratio,

which gives rise to a detectable chirp mass distribution much preferred by the data over a log-normal PBH model. An explicit lower and upper mass cutoff in these models also boosts their evidences significantly over the PBH scenario. While such high- and low-mass cutoffs might be expected in stellar-origin black holes, they are generally not expected in PBH models, which are consequently disfavored.

These statements hold true for almost any choices of the black hole mass function parameters and $f_{\rm PBH}$, although our modeling is expected to be inaccurate for high $f_{\rm PBH}$ and for very broad mass functions. We also studied simple extensions to the log-normal model in an attempt to better fit the data, finding that a bimodal mass function with small negative skewness (as expected from narrow peaks in the primordial power spectrum) provides a marginally improved fit, although at the cost of *a posteriori* reasoning and little physical motivation. At face value, our results strongly disfavor the possibility that all the sources seen in the first two LIGO-Virgo observing runs are merging PBH binaries forming from a smooth, symmetric peak in the primordial power spectrum. It is therefore worth discussing how these conclusions might be relaxed or challenged.

First, we caution that the empirical astrophysical models we have considered are really parametrizations and that we limit our analysis to the case that all the black holes are either astrophysical or primordial. A more sophisticated analysis would consider a mixed model with both forms of black holes. Our results are therefore not evidence against the possibility that LIGO-Virgo have detected PBHs but evidence against all of the detected black holes being PBHs.

To test a concrete example of a mixed stellar-PBH population of binary mergers, we computed the likelihoods of the PBH ($S = 1$) model and the LIGO models excluding the two sources possessing significant nonzero spin, GW151226 and GW170729. Since PBHs formed during radiation domination are expected to have zero spin at formation, these two sources are the most well motivated for exclusion from the analysis. Restricting to the reduced catalog of eight binary mergers, we find a slightly reduced best-fit $f_{\rm PBH} \approx 1.4 \times 10^{-3}$ and a smaller $\sigma \approx 0.48$, reflecting the fact that the two spinning sources lie near the extremes of the chirp mass distribution (GW151226 is the second-lightest system, and GW170729 is the heaviest). The maximum likelihood value of the PBH model compared to model B is slightly increased (-6.7, up from -7.1) with that of model A approximately unchanged. It thus appears that, even without the two spinning sources, we can expect the LIGO models to have significantly higher evidence compared with the PBH models, due to the PBH model struggling to fit the truncated and negatively skewed distribution of chirp masses.

Second, we have assumed that PBH binaries evolve from formation through to merger without modification to their dynamics from external astrophysical processes. In particular, we have neglected the possibility that matter accretes onto the binary, a potentially important effect influencing its angular momentum and mass [41,87]. Naively, one would expect accretion of material onto the binary to effectively skew the mass function toward heavier masses. As we have seen that the log-normal mass function already has too much positive skewness, this is likely to make the fit to the LIGO-Virgo data worse. It thus appears that the possibility that all sources are PBH binaries is even less likely with the inclusion of accretion.

Third, our results have sensitivity to the priors placed upon the model parameters. We have argued that there is little strong motivation for tightening the priors on the PBH model, but alternative choices could broaden them significantly. The abundance parameter $f_{\rm PBH}$ is exponentially sensitive to the amplitude of the peak in the primordial power spectrum which produced the PBHs. If we chose to impose a uniform prior on the order of magnitude of the peak amplitude, this would translate to a prior uniform in $\log \log f_{\rm PBH}$, which would potentially increase the prior volume of the PBH model by many orders of magnitude. However, this would only increase the degree to which the Bayes factor disfavors the PBH model.

Alternatively, we could boost the Bayes factor in favor of the PBH models by making alternative choices for the priors of the astrophysical models. We have argued that, if the functional form of this prior is left unmodified, this cannot result in the PBH model being favored over the astrophysical models without unphysical choices for the prior range. Alternatively, one could make alternative choices for the functional form of the astrophysical parameter priors. One can make the Bayes factor arbitrarily favorable toward the PBH model this way. It is possible that future models will find relations between the model parameters and more fundamental quantities related to the physics of stellar black hole binaries, in which case well-motivated choices for the priors of these fundamental parameters could result in a significantly broadened prior volume for the empirical parameters. It will therefore be necessary to rerun our analysis if such model refinements become available.

Is there a way to test the quality of the PBH model fit without reference to an empirical astrophysical model? Reference [21] found, using frequentist measures such as the χ^2 and KS test, that the LIGO-Virgo data are not an unlikely realization for a reasonable range of PBH model parameters. Our results are not in conflict with this conclusion, since we have focused largely on the *relative* quality of the model fits compared to astrophysically motivated models. We found that the detailed distribution of the chirp mass was the key discriminator in our tests. Since the χ^2 and KS tests do not make full use of this distribution but instead compress it down to test statistics, we do not expect these to be particularly powerful in

quantifying the quality of the PBH model fit. The Bayesian methodology has the advantage of using all the information available, which is one reason why we have focused on Bayesian model evidence ratios rather than frequentist statistical tests.

We have also neglected information coming from the spins of the merging black holes in our model comparisons. In reality, we expect the spin distributions of PBH and astrophysical mergers to be different, and including spin could impact our conclusions. A typical PBH spin distribution would have more weight in nonspinning objects, due to the negligible spin that PBHs are expected to have at formation. Given that all but two of the LIGO-Virgo sources we consider are consistent with zero spin, we might expect that including spin information boosts the relative probability of PBH models. Reference [32] performed a Bayesian comparison of PBH and astrophysical models using their differing predictions for black hole spin, finding that the data are not currently constraining enough to discriminate between the models. There is also theoretical uncertainty in how the initial spin distribution of PBHs evolves in the presence of accretion [41]. Furthermore, we have seen that excluding the two objects with nonzero spin has little effect on our conclusions. These considerations suggest that including spin in our analysis would not change our results significantly; the chirp mass, being a well-measured parameter for each system whose distribution is sensitive to model parameters, will likely remain the discriminating observable. A future extension of this work will be to include spin, with realistic astrophysical and PBH population distributions.

In this work, we have tested a specific model for the formation and subsequent evolution of a PBH binary. There exist models with dramatically different behavior allowing much larger $f_{\rm PBH}$, such as those of Refs. [15,53], which could yield quite different Bayesian evidences. However, there is no reason to expect that such models will provide good fits to the full set of GW events evidenced by the fact that the late-time capture model we studied does not provide a significantly different fit to the data. The subject of PBH binary evolution is an active and rapidly evolving field of study, and testing a broader range of merger models is a valuable extension of our formalism, which we defer to a future work.

It therefore appears that our conclusions are robust to including these added complications; both the full LIGO-Virgo sample of merging black holes and the subset consistent with zero spin are not well fit by PBH-PBH binaries compared with simple astrophysically motivated models.

Finally, it is interesting to consider how our results might change with the inclusion of recent new detections in the third observing run of LIGO-Virgo (O3). This has so far yielded a system, GW190814, with low mass-ratio $q \approx 0.1$ having one component in the lower mass gap with $m \approx 2.6\ M_\odot$ [45], the black hole binary merger GW190412 with low reported mass ratio $q \approx 0.3$ [44], the black hole merger GW190521 with total mass $150\ M_\odot$ and upper-mass gap component black holes [88], as well as roughly 50 new binary black hole detections. We caution that these three named sources have been singled out for publication by virtue of being "unusual," and hence the conclusions we can draw about population models are limited without including the full unbiased sample. A repeat of our analysis on the full sample of O3 events is forthcoming, so for now, we focus on these three unusual systems. First, we note that two low mass-ratio systems are more probable in PBH models than astrophysical models due to the extended mass function and lack of any mass correlation which might arise from mass transfer. As pointed out in Ref. [89], the constraint on the mass ratio of GW190412 is strongly dependent on the priors assumed for the source parameters, with $q \approx 1$ an equally good fit to the data when a low-spin prior (as might be expected for a PBH binary) is imposed. Focusing on chirp mass, which we have argued is where most of the constraining power comes from, GW190814 has $\mathcal{M}_{\rm chirp} = 6.09 \pm 0.06\ M_\odot$ and GW190412 has $\mathcal{M}_{\rm chirp} = 13.3 \pm 0.4\ M_\odot$, both in the source frame. Comparing the posterior predictive distribution given the O1 and O2 samples shows that both sources lie at the lighter end of the distribution, with GW190412 lying close to the peak of the PBH distribution. GW190814 is in the light tail of all the distributions we considered, although the constraining power of this object comes more from the implications of its low-mass component in the context of stellar black hole formation models. We note that GW190412 has a non-negligible spin parameter $\chi_{\rm eff} = 0.25^{+0.08}_{-0.11}$, while GW190814 is consistent with zero spin, raising the possibility that GW190412 is problematic for both PBH-PBH and stellar-stellar merger channel (although see the point above and Ref. [89]). GW190814, on the other hand, appears consistent with both, with high probability in the PBH model due to its low mass ratio. The source GW190521 has a chirp mass of roughly $65\ M_\odot$, with both components having non-negligible spin. Comparison with Fig. 11 shows that this source lies far in the high-mass tail of the distribution implied by the O1O2 sample. It is possible that the inclusion of this source in the sample could reduce some of the negative skewness in the chirp mass distribution and improve the PBH fit, although its chirp mass is so large that a worsened fit is also a possibility. Reference [90] discusses the possibility that GW190521 is a PBH binary, concluding that accretion is necessary to reconcile the implied merger rate with existing bounds on $f_{\rm PBH}$. A full analysis including all sources will shed more light on these intriguing issues.

With the sample size of black hole merger events expected to grow significantly with the conclusion of the third observing run of LIGO-Virgo and the newly online KAGRA facility [91], a principled statistical framework for analyzing the PBH merger scenario will prove increasingly

valuable. In this work, we have demonstrated the kind of analysis techniques that will be necessary to constrain the physics of primordial black holes in the coming era of gravitational wave astronomy.

ACKNOWLEDGMENTS

The authors thanks Nicola Bellomo, Christopher Berry, Gabriele Franciolini, Davide Gerosa, Karsten Jedamzik, Valerio de Luca, Chris Messenger, Paolo Pani, Antonio Riotto, and Ville Vaskonen for helpful correspondence. We thank the anonymous referees for helpful suggestions which improved the paper. A. H. acknowledges support from a Science and Technology Facilities Council Consolidated Grant. A. G. is funded by a Royal Society Studentship by means of a Royal Society Enhancement Award. C. B. acknowledges support from the Science and Technology Facilities Council (Grant No. ST/T000473/1). We acknowledge use of the software packages GWDET [35], DYNESTY [78], PyCBC [38], and LALSuite [92].

APPENDIX A: FAST NUMERICAL IMPLEMENTATION OF THE SUPPRESSION FACTOR

Bayesian inference of the posterior for PBH mergers involves many computations of the likelihood, and fast implementations are therefore crucial. The computational bottleneck in the likelihood evaluation step is the computation of the suppression factor, given in Eq. (9) as

$$S = \frac{e^{-\bar{N}(y)}}{\Gamma(21/37)} \int_0^\infty \mathrm{d}v\, v^{-\frac{16}{37}} \exp\left[-\bar{N}(y)\langle m\rangle \int_0^\infty \frac{\mathrm{d}m}{m}\psi(m) F\left(\frac{m}{\langle m\rangle}\frac{v}{\bar{N}(y)}\right) - \frac{3\sigma_M^2 v^2}{10 f_{\mathrm{PBH}}^2}\right], \tag{A1}$$

where we remind the reader that $F(x) = {}_1F_2(-1/2; 3/4, 5/4; -9x^2/16) - 1$ with ${}_1F_2$ a generalized hypergeometric function.

Fast evaluation of Eq. (A1) is numerically challenging for general mass functions due to the double integral and potential for large dynamic range in $\psi(m)$.[11] We can make progress, however, by using the Taylor expansion of $F(x)$ at both high and low values of its argument.

1. Switching approximation to $F(x)$

Around $z = 0$, we can truncate the definition of the generalized hypergeometric function to obtain the Taylor series of ${}_1F_2$ to order $n_{\max}$,

$${}_1F_2(a_1; b_2, b_2; z) \approx \sum_{n=0}^{n_{\max}} \frac{(a_1)_n}{(b_1)_n (b_2)_n} \frac{z^n}{n!}, \tag{A2}$$

where $(\alpha)_n$ is a Pochhammer symbol given by

$$(\alpha)_n = \alpha(\alpha+1)(\alpha+2)\ldots(\alpha+n-1) \tag{A3}$$

for $n \geq 1$, with $(\alpha)_0 = 1$. We choose $n_{\max} = 3$, additionally computing the next-order $n = 4$ term to quantify the perturbative error.

At large values of z, we use the asymptotic expansion available from Ref. [93], given by

$$\begin{aligned}
{}_1F_2(a_1; b_1, b_2; z) \approx{}& \frac{\Gamma(b_1)\Gamma(b_2)}{\Gamma(b_1-a_1)\Gamma(b_2-a_2)} (-z)^{-a_1} \left\{1 + \frac{a_1(a_1-b_1+1)(a_1-b_2+1)}{z}\right. \\
&\left. + \frac{1}{2z^2}[a_1(a_1+1)(a_1-b_1+1)(a_1-b_1+2)(a_1-b_2+2)(a_1-b_2+1)] + \ldots\right\} \\
&+ \frac{\Gamma(b_1)\Gamma(b_2)}{2\sqrt{\pi}\Gamma(a_1)} (-z)^{\frac{1}{2}(a_1-b_1-b_2+\frac{1}{2})} \\
&\times \left\{e^{-i[\frac{1}{2}\pi(a_1-b_1-b_2+\frac{1}{2})+2\sqrt{-z}]}\left(1 + \frac{d_1}{\sqrt{-z}} + \frac{d_2}{z} + \ldots\right) + e^{i[\frac{1}{2}\pi(a_1-b_1-b_2+\frac{1}{2})+2\sqrt{-z}]}\left(1 - \frac{d_1}{\sqrt{-z}} + \frac{d_2}{z} + \ldots\right)\right\},
\end{aligned} \tag{A4}$$

[11] Additionally, recently, there was no public implementation of the ${}_1F_2$ hypergeometric function available in PYTHON.

where

$$d_1 = \frac{i}{16}[-3 + 12a_1^2 - 4b_1^2 + 8b_2 - 4b_2^2 + 8b_1(1+b_2) - 8a_1(1+b_1+b_2)], \tag{A5}$$

$$\begin{aligned} d_2 = \frac{1}{512}\{&-15 + 144a_1^4 + 16b_1^4 + 16b_2 + 56b_2^2 - 64b_2^3 + 16b_2^4 - 64b_1^3(1+b_2) - 64a_1^3(7+3b_1+3b_2) \\ &+ 8b_1^2(7+8b_2+12b_2^2) + 16b_1(1+25b_2+4b_2^2-4b_2^3) - 8a_1^2[-43+4b_1^2-72b_2+4b_2^2-8b_1(9+5b_2)] \\ &+ 16a_1[-1+4b_1^3-25b_2-4b_2^2+4b_2^3-4b_1^2(1+b_2)-b_1(25+40b_2+4b_2^2)]\}. \end{aligned} \tag{A6}$$

The complex exponentials in this expansion give rise to oscillating terms in $F(x)$ at order $\mathcal{O}(x^{-n})$ for $n \geq 2$. Neglecting these terms, $F(x)$ has the asymptotic expansion for $x \gg 1$:

$$F(x) \approx x - 1 + \frac{1}{6x} + \ldots \tag{A7}$$

At small values of x, we use Eq. (A2), keeping terms in ${}_1F_2$ up to order z^3 [i.e., terms of order x^6 in $F(x)$]. We additionally compute the error arising from neglect of the x^8 term. At some value x_*, we switch to the asymptotic expansion of Eq. (A4), keeping terms up to order x^{-3} and using the neglected x^{-4} term to approximate the error (with oscillating terms set to their maximal value). The switching value x_* is then chosen to minimize the relative error of the approximation. This yields $x_* = 2.72$ in the case when all x^{-4} terms are used and $x_* = 2.74$ when the approximation equation (A7) is used.

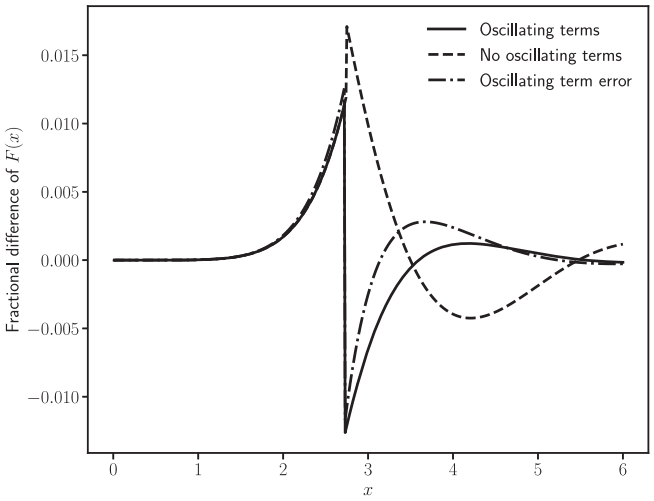

FIG. 16. Fractional difference of the approximation described in the text to the function $F(x)$, defined after Eq. (A1), and its true value with (solid) and without (dashed) oscillating terms. We also show the estimate of the error of the approximation based on the neglected higher-order terms (dot-dashed). We achieve better than 2% accuracy across all values of the argument and additionally can accurately predict the error of the approximation. Note that all curves are indistinguishable for small values of x.

In Fig. 16, we show the relative error of our approximation scheme compared to an exact calculation implemented in MATLAB. The errors are generally subpercent, reaching maximal values of less than or approximately equal to 2% around x_*. The error estimate derived from neglected higher-order terms is generally an accurate approximation to the true error, and we also see that oscillating terms are generally negligible with Eq. (A7) sufficient for $x \gtrsim x_*$. The maximum absolute error of our approximation is also approximately $\mathcal{O}(1\%)$.

The results of this section may be summarized as follows: in our baseline PBH likelihood analysis, we use the following approximation for $F(x)$, accurate at the percent level:

$$F(x) \approx \begin{cases} \frac{3}{10}x^2 - \frac{3}{280}x^4 + \frac{27}{80080}x^6 & x \leq 2.74 \\ x - 1 + \frac{1}{6x} & x > 2.74. \end{cases} \tag{A8}$$

2. Exact integration for log-normal mass functions

A polynomial expansion for $F(x)$ is particularly useful when $\psi(m)$ is a log-normal distribution, since the innermost integral in Eq. (A1) can be done analytically term by term. For this, we need to neglect the small oscillating terms in the large-x expansion of $F(x)$, which by Fig. 16 are negligible at the percent level. We also need the results

$$\begin{aligned} I_p^+(m_c,\sigma;m_*) &\equiv \int_0^{m_*} \mathrm{d}m \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\ln^2(m/m_c)}{2\sigma^2}\right] m^{p-2} \\ &= m_c^{p-1} e^{\frac{(1-p)^2\sigma^2}{2}} \Phi\left[\frac{1}{\sigma}\ln(m_*/m_c) + (1-p)\sigma\right], \end{aligned} \tag{A9}$$

$$\begin{aligned} &I_p^-(m_c,\sigma;m_*) \\ &\equiv \int_{m_*}^{\infty} \mathrm{d}m \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\ln^2(m/m_c)}{2\sigma^2}\right] m^{p-2} \\ &= m_c^{p-1} e^{\frac{(1-p)^2\sigma^2}{2}} \left\{1 - \Phi\left[\frac{1}{\sigma}\ln(m_*/m_c) + (1-p)\sigma\right]\right\}, \end{aligned} \tag{A10}$$

where $\Phi(x)$ is the cumulative distribution function of the normal distribution (expressible in terms of the error function). Substituting in the polynomial expansions for $F(x)$ at small x and large x, we can use Eq. (A10) to leave only the outermost integral in Eq. (A1). We perform this integral using numerical quadrature, which results in a fast and accurate (to roughly 1%) approximation to the suppression factor which can be used in likelihood evaluations.

Using that $\langle m \rangle = m_c e^{-\sigma^2/2}$ for a log-normal mass function,[12] it is straightforward to verify that at fixed $\bar{N}(y)$ the suppression factor is independent of the absolute mass scale m_c and depends only on the width σ. Physically, this is due to the assumption that the PBHs are distributed in space in a way that is independent of their mass. In reality, this assumption might be broken if PBHs cluster significantly due to mass segregation, a complication which we neglect here.

It is useful to consider a few limiting cases of the suppression factor. First, for any mass function, in the limit that $F(x)$ is dominated by its leading-order quadratic part at small x, the distribution of angular momentum tends to a Gaussian, and the suppression factor is [17]

$$S_{\min} = \frac{\pi}{\Gamma(29/37)} \left(\frac{\sigma_j}{j_0}\right)^{-\frac{21}{37}} e^{-\bar{N}(y)}, \tag{A11}$$

where

$$\frac{\sigma_j^2}{j_0^2} = \frac{6}{5}\left[\frac{1+\sigma_m^2/\langle m\rangle^2}{\bar{N}(y)} + \frac{\sigma_M^2}{f_{\rm PBH}^2}\right], \tag{A12}$$

and $\sigma_m \equiv \sqrt{\langle m^2\rangle - \langle m\rangle^2}$ is the width of the mass function. It may be shown that $S \geq S_{\min}$. Physically, this limit is realized when $\bar{N}(y) \to \infty$; i.e., the spatial density of PBHs becomes sufficiently large that the central limit theorem Gaussianizes the distribution of torquing angular momenta (which has already been assumed to have happened for the dark matter component). The variance of this Gaussian is σ_j^2, which has contributions from the PBHs and dark matter adding in quadrature. The model of Ref. [17] imposes an exponential suppression in this regime, realized via the $e^{-\bar{N}(y)}$ term in Eq. (A11), which accounts for the fact that a high spatial density of PBHs will increase the likelihood of a PBH being sufficiently close to the binary as to prevent its formation due to three-body effects. In our implementation, we simply set the suppression factor to zero for $\bar{N}(y) > 23$, where the exponential suppression ensures that S is practically zero.

In the opposite limit, the term linear in x dominates in $F(x)$, and the suppression factor tends to

$$S_{\max} = \left(\frac{5f_{\rm PBH}^2}{6\sigma_M^2}\right)^{\frac{21}{74}} U\left(\frac{21}{74}, \frac{1}{2}, \frac{5f_{\rm PBH}^2}{6\sigma_M^2}\right), \tag{A13}$$

where U is a confluent hypergeometric function. One can show that $S \leq S_{\max}$. This limit is realized when $\bar{N}(y) \to 0$, although how quickly the limit is reached depends on $f_{\rm PBH}/\sigma_M$. Reference [17] advocates $\bar{N}(y) \ll f_{\rm PBH}^2/\sigma_M^2$. We find that Eq. (A13) is accurate to less than or approximately equal to 2% for all $f_{\rm PBH} > 10^{-6}$ when $\bar{N}(y) < 0.01$ and for all σ we consider. We adopt this threshold for $\bar{N}(y)$ when switching to the asymptotic limit (A13). This gives percent-level accuracy for the merger rate while maintaining a reasonable run time.

APPENDIX B: SUPPRESSION-INDUCED PARAMETER DEGENERACIES IN THE PBH MODEL

When sampling from the posterior in the PBH model, we find a pronounced degeneracy tail in each of the three projected two-dimensional planes defined by $f_{\rm PBH}$, m_c, and σ; see Fig. 7. The feature allows for large values of all three parameters and is not present when the model is analyzed fixing the suppression factor to unity. In this section, we discuss the origin of the degeneracy feature.

In Fig. 17, we show posterior samples in the parameter space $(\log_{10} m_c, \sigma^2)$ color coded by the value of $f_{\rm PBH}$. The degeneracy feature appears linear in this space and roughly corresponds to fixing $\langle m \rangle = m_c e^{-\sigma^2/2} \approx 20\ M_\odot$, i.e., fixing the average PBH mass to the observed mass scale of the LIGO sources.

In the middle and right panels of Fig. 17, we split the samples into a set with $10^{-2} < f_{\rm PBH} < 10^{-1}$ and a set with $10^{-4} < f_{\rm PBH} < 10^{-2}$, respectively. Samples in the degeneracy tail typically have higher $f_{\rm PBH}$, which indicates that we are looking at the projection of a three-parameter degeneracy.

In Fig. 18, we split the samples by the degree of suppression, defined as the ratio between the total detectable number of mergers β in a model with suppression factor set to unity to its value in a model with suppression set by Eqs. (9) and (7). The degeneracy tail is clearly characterized by higher degrees of suppression, which compensate for the overproduction of mergers that large values of $f_{\rm PBH}$ would otherwise imply.

These plots suggest that the degeneracy tail is caused by the dependence of the suppression factor on the mass function parameters via Eq. (7), namely,

$$\bar{N}(y) = \frac{M}{\langle m\rangle} \frac{f_{\rm PBH}}{f_{\rm PBH} + \sigma_M}, \tag{B1}$$

where $\sigma_M \approx 0.006$ and M is the total binary mass. We remind the reader that $\bar{N}(y)$ is the expected number of

[12] We remind the reader that angle brackets denote expectation values over $dn/dm \propto \psi/m$, i.e., $\langle m\rangle \equiv [\int dm m\psi(m)/m]/\int dm\psi(m)/m$.

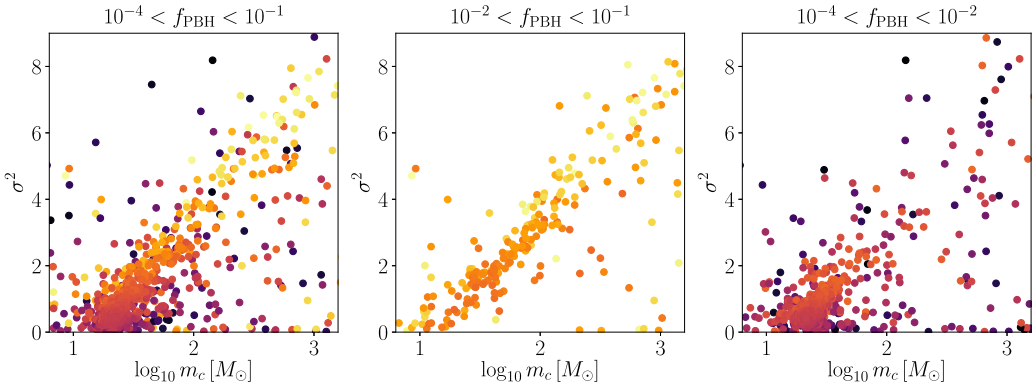

FIG. 17. Posterior samples in the $(\log_{10} m_c, \sigma^2)$ plane for the suppression factor PBH model and GWTC-1 data, color coded by $f_{\rm PBH}$ (warmer color corresponding to larger values). Left panel: samples having $10^{-4} < f_{\rm PBH} < 10^{-1}$, thinned by a factor of 4 for visual clarity. Middle panel: samples with higher values of $f_{\rm PBH}$ in the range $10^{-2} < f_{\rm PBH} < 10^{-1}$. Right panel: samples with lower values of $f_{\rm PBH}$ in the range $10^{-4} < f_{\rm PBH} < 10^{-2}$. Note that the color coding is the same across the three panels.

PBHs in a spherical region of comoving radius y which contains no other black holes except the binary pair and that the suppression factor has a prefactor of $e^{-\bar{N}(y)}$, introduced in Ref. [17] to ensure that no other PBH gets close enough to the binary to disrupt it prior to the merger event.

The suppression factor enters the likelihood in two places: in the average over source parameter MCMC samples [the term in square brackets in Eq. (15)] and in the $e^{-\beta}$ factor entering via the Poisson probability of seeing $N_{\rm obs}$ events given β were expected. The former essentially fixes the total mass in Eq. (B1) to the LIGO mass scale, which implies that $\langle m \rangle$ is fixed to keep the source likelihoods high and unsuppressed; note that in the degeneracy tail $f_{\rm PBH} \gg \sigma_M$ such that the second term in Eq. (B1) is irrelevant. Models with high σ have contributions to β from a much broader range of masses; the typical total masses contributing to β [Eq. (11)] are approximately $2m_c$, meaning $M \gg \langle m \rangle$ when σ is large. This implies that $\bar{N}(y) \gg 1$ for much of the integration range in β, which in turn implies a high degree of suppression in the total number of mergers and compensates for the high $f_{\rm PBH}$.

We note that the degeneracy tail skews the one-dimensional posteriors on the PBH model parameters but has a relatively minor impact on their median and best-fit values. Nevertheless, what we have uncovered is a mechanism for generating high values of $f_{\rm PBH}$ which can still fit the LIGO data. Equation (B1) was arrived at in Ref. [17] by a combination of simulation results and analytic arguments, but since it was not tested for the extreme mass function parameters favored in the degeneracy feature, the model could well be unreliable in this regime. Further investigations with N-body simulations will be required to rigorously test the degeneracy feature seen in Fig. 7.

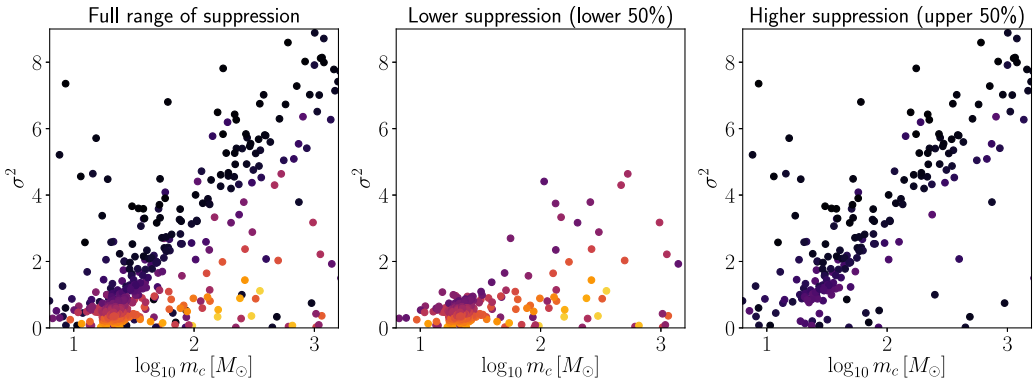

FIG. 18. Posterior samples in the $(\log_{10} m_c, \sigma^2)$ plane for the suppression factor PBH model and GWTC-1 data, color coded by suppression factor, defined as $\beta(S=1)/\beta(S)$ (warmer color corresponding to lower suppression). Left panel: sll samples, thinned by a factor of 8 for visual clarity. Middle panel: samples with suppression in the lower 50% quantile. Right panel: samples with suppression in the upper 50% quantile. Note that the color coding is the same across the three panels.

APPENDIX C: THE POSTERIOR PREDICTIVE DISTRIBUTION

In this section, we present a derivation of the PPD. This quantity is sufficient to understand the differences in Bayesian evidences between our BBH merger models.

Each model has the flexibility to fit the total number of observed mergers by adjusting an amplitudelike parameter ($f_{\rm PBH}$ or R_0), so the best-fitting β is roughly 10 for all models. The factor of $e^{-\beta}$ in the likelihood equation (13) thus cancels in the likelihood ratio and evidence ratio, and we can approximate D with $\mathbf{d}_{\rm new}$, the new GW strain data. Formally, the likelihood for $\mathbf{d}$ is that used to produce the posterior samples for the source parameters. We will instead simplify the analysis by compressing the data to a set of estimators for the source parameters which best constrain the population models. For both the PBH and LIGO models, the relevant source parameters are the total mass M, the mass ratio q, and the redshift z, or combinations of these. We have seen from Fig. 10 that the redshift distributions of the observable mergers in both models are indistinguishable, so we can anticipate that the likelihood in M and q will be sufficient to explain the differing maximum likelihood values of the models. Since it is the detector-frame chirp mass $\mathcal{M}_z$ which is most closely related to the signal measured in the data (as discussed in Sec. II), our two estimators will be $\widehat{\mathcal{M}}_z$ and $\hat{q}$, where the precise expression of these in terms of the data is left unspecified for now.

Although the preferred q-distributions are quite different (middle panel of Fig. 10), the typical errors on q are large for the GWTC-1 sources, so it is unclear how powerful the mass ratio distribution is in discriminating between models. To test this, we generated new source parameter posteriors for each merger event by randomizing the q values at each sample point. Specifically, for each sample $\boldsymbol{\lambda}_i$, we recorded the chirp mass $\mathcal{M}_{\rm chirp}$, replaced q with a random sample from a uniform distribution between 0 and 1, and then set m_1 and m_2 using the saved $\mathcal{M}_{\rm chirp}$ and the new q. We then reran the inference of the PBH $S = 1$ model and the LIGO models A and B using these new source posteriors. This procedure destroys all information on mass ratio in each source, preserving that on redshift and detector-frame chirp mass. The (natural log) evidence ratio of the PBH model to model B with this new data is -7.30 ± 0.25. For the PBH model to model A, the evidence ratio is -5.72 ± 0.23, and for model A to model B, it is -1.59 ± 0.23. The $S = 1$ PBH model is now slightly less disfavored compared with model A and model B but is still heavily disfavored compared with both these models. There is no significant change in the evidence ratio between the two LIGO models. This test strongly suggests that the distribution of chirp masses preferred by the models is the key discriminator between them. The mass ratio uncertainties are too large in the GWTC-1 catalog for q to be effective at constraining the space of allowed models.

These arguments strongly suggest that the source likelihood $p(\mathbf{d}|\boldsymbol{\theta}, M)$ needed for the PPD should be the probability of an "observed chirp mass" $\widehat{\mathcal{M}}_z$ given source parameters. An expression for $p(\mathbf{d}|\boldsymbol{\theta}, M)$ is given in Ref. [71]. For a single source with *detectable* GW strain $\mathbf{d}$ and source parameters $\boldsymbol{\lambda}$, the likelihood is

$$p(\mathbf{d}|\boldsymbol{\theta}) = \frac{I(\mathbf{d}) \int \mathrm{d}\boldsymbol{\lambda} p(\mathbf{d}|\boldsymbol{\lambda}) p(\boldsymbol{\lambda}|\boldsymbol{\theta})}{\int \mathrm{d}\boldsymbol{\lambda} p_{\rm det}(\boldsymbol{\lambda}) p(\boldsymbol{\lambda}|\boldsymbol{\theta})}, \tag{C1}$$

where $I(\mathbf{d})$ is unity if the data pass the detection threshold and zero if it does not. We have also defined the detection probability over the complete set of source parameters $p_{\rm det}(\boldsymbol{\lambda})$ as

$$p_{\rm det}(\boldsymbol{\lambda}) = \int \mathrm{d}\mathbf{d} I(\mathbf{d}) p(\mathbf{d}|\boldsymbol{\lambda}), \tag{C2}$$

where $p(\mathbf{d}|\boldsymbol{\lambda})$ is the probability of any dataset, not just those observable by the detector. The joint likelihood of N sources is simply the product of individual likelihoods each given by Eq. (C1). The prior distribution of source parameters given the population model $p(\boldsymbol{\lambda}|\boldsymbol{\theta})$ is simply proportional to the merger rate $\mathrm{d}N/\mathrm{d}m_1\mathrm{d}m_2\mathrm{d}z$ appearing in Eq. (13). Note that the overall normalization of the merger rate drops out of Eq. (C1).

Equation (C1) is the likelihood for GW strain data, and we wish to rewrite it as a one-dimensional likelihood for observed chirp mass $\widehat{\mathcal{M}}_z$. This will allow us to study the origin of the large likelihood ratios between models, which we have argued is primarily due to the relative ability of models to fit the observed distribution of chirp masses. It will also allow an assessment of the absolute quality of model fits via the PPD. To do this, we will make a series of well-motivated approximations to construct a likelihood for the compressed data $\widehat{\mathcal{M}}_z$.

First, we will assume that $I(\mathbf{d}) = \Theta(\hat{\rho} - \rho_*)$ where Θ is the Heaviside step function, $\hat{\rho}$ is the S/N of the observed waveform, and ρ_* is a threshold S/N; i.e., we assume the merger is detectable if its S/N is above a sharp threshold. Second, we assume that $\hat{\rho}$ is unaffected by noise fluctuations, such that for source parameters $\boldsymbol{\lambda}$ we have $p(\hat{\rho}|\boldsymbol{\lambda}) = \delta^D[\hat{\rho} - \rho(\boldsymbol{\lambda})]$, where δ^D is the Dirac delta function and $\rho(\boldsymbol{\lambda})$ is the S/N of a model template with source parameters $\boldsymbol{\lambda}$. This is a reasonable approximation since $\hat{\rho}$ is a stack across the whole waveform and is typically well constrained. These two approximations imply that $p_{\rm det}(\boldsymbol{\lambda}) = \Theta[\rho(\boldsymbol{\lambda}) - \rho_*]$. Since $\hat{\rho}$ is assumed to be nonstochastic, we can integrate it out of Eq. (C1) and redefine the data vector $\mathbf{d}$ as having the overall amplitude projected out.

Next, we assume that the S/N can be written in terms of the orientation parameter ω introduced in Sec. II as

$\rho(\boldsymbol{\lambda}) = \omega\rho_{\rm opt}(\tilde{\boldsymbol{\lambda}})$, where $\tilde{\boldsymbol{\lambda}}$ are the source parameters excluding the orientation and angular position parameters (which have been combined into ω) and $\rho_{\rm opt}$ is the S/N of an optimally oriented binary. For isotropic sources, we have $p(\boldsymbol{\lambda}|\boldsymbol{\theta}) = p(\tilde{\boldsymbol{\lambda}}|\boldsymbol{\theta})p(\omega)$, where $p(\omega)$ is the distribution discussed in Sec. II and produced by GWDET. With these approximations, Eq. (C1) becomes

$$p(\mathbf{d}|\boldsymbol{\theta}) \propto \int \mathrm{d}\tilde{\boldsymbol{\lambda}}\mathrm{d}\omega\Theta[\omega\rho_{\rm opt}(\tilde{\boldsymbol{\lambda}}) - \rho_*]p(\mathbf{d}|\tilde{\boldsymbol{\lambda}},\omega)p(\omega)p(\tilde{\boldsymbol{\lambda}}|\boldsymbol{\theta}) \propto \int \mathrm{d}\tilde{\boldsymbol{\lambda}}p(\tilde{\boldsymbol{\lambda}}|\boldsymbol{\theta})\int_{\frac{\rho_*}{\rho_{\rm opt}(\tilde{\boldsymbol{\lambda}})}}^{1}\mathrm{d}\omega p(\omega)p(\mathbf{d}|\tilde{\boldsymbol{\lambda}},\omega). \tag{C3}$$

Now, we will assume that an estimator $\widehat{\mathcal{M}}_z$ for the chirp mass can be constructed from the data using some form of massive data compression. Writing $\mathbf{d} = (\widehat{\mathcal{M}}_z, \tilde{\mathbf{d}})$, we can integrate out all other "modes" of the data $\tilde{\mathbf{d}}$ which keep $\widehat{\mathcal{M}}_z$ fixed. This simply amounts to replacing $\mathbf{d}$ with $\widehat{\mathcal{M}}_z$ everywhere in Eq. (C3).

Our next approximation sets $p(\widehat{\mathcal{M}}_z|\tilde{\boldsymbol{\lambda}},\omega) \approx p(\widehat{\mathcal{M}}_z|\tilde{\boldsymbol{\lambda}})$; i.e., we assume the estimated chirp mass on its own provides no information on the orientation or angular position of the binary. With this, we have

$$p(\widehat{\mathcal{M}}_z|\boldsymbol{\theta}) \propto \int \mathrm{d}\tilde{\boldsymbol{\lambda}}p_{\rm det}(\tilde{\boldsymbol{\lambda}})p(\widehat{\mathcal{M}}_z|\tilde{\boldsymbol{\lambda}})p(\tilde{\boldsymbol{\lambda}}|\boldsymbol{\theta}), \tag{C4}$$

where we used Eq. (1) to substitute for the angle-averaged detection probability.

We now assume that both $p_{\rm det}(\tilde{\boldsymbol{\lambda}})$ and $p(\widehat{\mathcal{M}}_z|\tilde{\boldsymbol{\lambda}})$ depend only upon $\mathcal{M}_z$, q, and z; i.e., we neglect any spin dependence in the S/N. This allows us to integrate out all other source parameters from Eq. (C4). This just leaves $p(\widehat{\mathcal{M}}_z|\mathcal{M}_z,q,z)$, the likelihood for the observed chirp mass, to be specified.

At this point, we must make a choice for the form of $p(\widehat{\mathcal{M}}_z|\mathcal{M}_z,q,z)$, since so far we have only used it as a function of model parameters and not of "observed" parameters. Since the data compression producing $\widehat{\mathcal{M}}_z$ is massive, we will assume that the sampling distribution of the estimator is Gaussian with mean μ and variance σ^2. An unbiased estimator for the chirp mass would have $\mu = \mathcal{M}_z$, but the variance can have a general dependence on $\mathcal{M}_z$, q, and z. This could in principle be inferred from the source parameter posterior given priors on the source parameters, but we will make the ansatz that σ^2 is roughly constant and can be set equal to the marginalized *posterior* variances on chirp mass from the MCMC samples—this is justified if $p(\widehat{\mathcal{M}}_z|\mathcal{M}_z)$ really is Gaussian and the prior on chirp mass is uniform (although recall that the chirp mass is measured with approximately 15% precision with little sensitivity to the choice of prior).

With these assumptions, we can write

$$p(\widehat{\mathcal{M}}_z|\boldsymbol{\theta}) = \int \mathrm{d}\mathcal{M}_z p(\widehat{\mathcal{M}}_z|\mathcal{M}_z)p(\mathcal{M}_z|\boldsymbol{\theta}), \tag{C5}$$

$$p(\mathcal{M}_z|\boldsymbol{\theta}) = \frac{\int \mathrm{d}q\mathrm{d}z p_{\rm det}(\mathcal{M}_z,q,z)p(\mathcal{M}_z,q,z|\boldsymbol{\theta})}{\int \mathrm{d}\mathcal{M}_z\mathrm{d}q\mathrm{d}z p_{\rm det}(\mathcal{M}_z,q,z)p(\mathcal{M}_z,q,z|\boldsymbol{\theta})}. \tag{C6}$$

Equation (C5) is simply the predicted distribution of chirp mass averaged over all other source parameters *convolved* with the observational uncertainty specified by $p(\widehat{\mathcal{M}}_z|\mathcal{M}_z)$. We immediately recognize Eq. (C6) as a normalized version of the differential detectable merger rate, i.e., $\partial \ln\beta/\partial\mathcal{M}_z$. Recall that in Fig. 10 we plotted $\partial\beta/\partial M$, $\partial\beta/\partial q$, and $\partial\beta/\partial z$.

Finally, we can average over the posterior of the model parameters $\boldsymbol{\theta}$ for each source to get the PPD. This gives, for the full catalog of sources,

$$p(\{\widehat{\mathcal{M}}_z\}|\{\mathbf{d}\}) = \prod_{i=1}^{N_{\rm obs}}\int \mathrm{d}\mathcal{M}_z p(\widehat{\mathcal{M}}_z^i|\mathcal{M}_z)p(\mathcal{M}_z|\{\mathbf{d}\}), \tag{C7}$$

$$p(\mathcal{M}_z|\{\mathbf{d}\}) = \int \mathrm{d}\boldsymbol{\theta}p(\mathcal{M}_z|\boldsymbol{\theta})p(\boldsymbol{\theta}|\{\mathbf{d}\}). \tag{C8}$$

In a slight abuse of terminology, we will also refer to $p(\mathcal{M}_z|\{\mathbf{d}\})$ as the PPD. This object quantifies the probability distribution of the detector-frame chirp mass given the data and can be convolved with the observational errors according to Eq. (C7) to give an equivalent predictive distribution for the observed chirp mass.

It is straightforward to show that Eq. (C5) evaluated at the locations of the observed strain data gives a likelihood function equivalent to that specified in Eq. (13) and used throughout this work for inference. Indeed, the above derivation shows how this simplified likelihood may be obtained from first principles and makes all approximations transparent. The only difference with Eq. (13) is the Poisson probability of observing $N_{\rm obs}$ sources when β were expected—the likelihood equation (C5) only accounts for the relative merger rate, which is sufficient to understand the disparity in likelihood ratios since $\beta \approx 10$ at the best-fit point for all models.

[1] Y. B. Zel'dovich and I. D. Novikov, Sov. Astron. **10**, 602 (1967).
[2] S. Hawking, Mon. Not. R. Astron. Soc. **152**, 75 (1971).
[3] B. J. Carr and S. W. Hawking, Mon. Not. R. Astron. Soc. **168**, 399 (1974).
[4] B. J. Carr, Astrophys. J. **201**, 1 (1975).
[5] B. Carr, F. Kühnel, and M. Sandstad, Phys. Rev. D **94**, 083504 (2016).
[6] B. Carr and F. Kühnel, Annu. Rev. Nucl. Part. Sci. **70**, 355 (2020).
[7] B. Carr, K. Kohri, Y. Sendouda, and J. Yokoyama, arXiv:2002.12778.
[8] A. M. Green and B. J. Kavanagh, arXiv:2007.10722.
[9] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116**, 061102 (2016).
[10] S. Bird, I. Cholis, J. B. Muñoz, Y. Ali-Haïmoud, M. Kamionkowski, E. D. Kovetz, A. Raccanelli, and A. G. Riess, Phys. Rev. Lett. **116**, 201301 (2016).
[11] M. Sasaki, T. Suyama, T. Tanaka, and S. Yokoyama, Phys. Rev. Lett. **117**, 061101 (2016).
[12] S. Clesse and J. García-Bellido, Phys. Dark Universe **15**, 142 (2017).
[13] M. Sasaki, T. Suyama, T. Tanaka, and S. Yokoyama, Classical Quantum Gravity **35**, 063001 (2018).
[14] V. De Luca, G. Franciolini, P. Pani, and A. Riotto, J. Cosmol. Astropart. Phys. 06 (2020) 044.
[15] K. Jedamzik, J. Cosmol. Astropart. Phys. 09 (2020) 022.
[16] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. X **9**, 031040 (2019).
[17] M. Raidal, V. Vaskonen, and H. Veermäe, J. Cosmol. Astropart. Phys. 09 (2017) 037.
[18] Z.-C. Chen and Q.-G. Huang, Astrophys. J. **864**, 61 (2018).
[19] Y. Wu, Phys. Rev. D **101**, 083008 (2020).
[20] Y. Ali-Haïmoud, E. D. Kovetz, and M. Kamionkowski, Phys. Rev. D **96**, 123523 (2017).
[21] A. D. Gow, C. T. Byrnes, A. Hall, and J. A. Peacock, J. Cosmol. Astropart. Phys. 01 (2020) 031.
[22] Y.-F. Cai, X. Tong, D.-G. Wang, and S.-F. Yan, Phys. Rev. Lett. **121**, 081306 (2018).
[23] C. T. Byrnes, M. Hindmarsh, S. Young, and M. R. S. Hawkins, J. Cosmol. Astropart. Phys. 08 (2018) 041.
[24] V. Vaskonen and H. Veermäe, Phys. Rev. D **101**, 043015 (2020).
[25] A. D. Gow, C. T. Byrnes, P. S. Cole, and S. Young, arXiv:2008.03289.
[26] M. Raidal, C. Spethmann, V. Vaskonen, and H. Veermäe, J. Cosmol. Astropart. Phys. 02 (2019) 018.
[27] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaboration), Phys. Rev. X **6**, 041015 (2016).
[28] M. Fishbach and D. E. Holz, Astrophys. J. Lett. **851**, L25 (2017).
[29] M. Fishbach, D. E. Holz, and W. M. Farr, Astrophys. J. Lett. **863**, L41 (2018).
[30] D. Gerosa, E. Berti, R. O'Shaughnessy, K. Belczynski, M. Kesden, D. Wysocki, and W. Gladysz, Phys. Rev. D **98**, 084036 (2018).
[31] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaboration), Astrophys. J. Lett. **882**, L24 (2019).
[32] N. Fernandez and S. Profumo, J. Cosmol. Astropart. Phys. 08 (2019) 022.
[33] P. A. R. Ade *et al.* (Planck Collaboration), Astron. Astrophys. **594**, A13 (2016).
[34] A. D. Dolgov, A. G. Kuranov, N. A. Mitichkin, S. Porey, K. A. Postnov, O. S. Sazhina, and I. V. Simkin, arXiv:2005.00892.
[35] D. Gerosa, Dgerosa/Gwdet: V0.1, 2017.
[36] L. S. Finn and D. F. Chernoff, Phys. Rev. D **47**, 2198 (1993).
[37] D. Gerosa, S. Ma, K. W. K. Wong, E. Berti, R. O'Shaughnessy, Y. Chen, and K. Belczynski, Phys. Rev. D **99**, 103004 (2019).
[38] S. A. Usman, A. H. Nitz, I. W. Harry, C. M. Biwer, D. A. Brown, M. Cabero, C. D. Capano, T. Dal Canton, T. Dent, S. Fairhurst *et al.*, Classical Quantum Gravity **33**, 215004 (2016).
[39] V. De Luca, V. Desjacques, G. Franciolini, A. Malhotra, and A. Riotto, J. Cosmol. Astropart. Phys. 05 (2019) 018.
[40] M. Mirbabayi, A. Gruzinov, and J. Noreña, J. Cosmol. Astropart. Phys. 03 (2020) 017.
[41] V. De Luca, G. Franciolini, P. Pani, and A. Riotto, J. Cosmol. Astropart. Phys. 04 (2020) 052.
[42] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaboration), Living Rev. Relativity **19**, 1 (2016).
[43] D. Wysocki, J. Lange, and R. O'Shaughnessy, Phys. Rev. D **100**, 043012 (2019).
[44] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. D **102**, 043015 (2020).
[45] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Astrophys. J. Lett. **896**, L44 (2020).
[46] J. Roulet and M. Zaldarriaga, Mon. Not. R. Astron. Soc. **484**, 4216 (2019).
[47] T. Nakama, J. Silk, and M. Kamionkowski, Phys. Rev. D **95**, 043511 (2017).
[48] B. J. Kavanagh, D. Gaggero, and G. Bertone, Phys. Rev. D **98**, 023536 (2018).
[49] P. C. Peters, Phys. Rev. **136**, B1224 (1964).
[50] Y. N. Eroshenko, J. Phys. Conf. Ser. **1051**, 012010 (2018).
[51] K. Ioka, T. Chiba, T. Tanaka, and T. Nakamura, Phys. Rev. D **58**, 063003 (1998).
[52] D. Inman and Y. Ali-Haïmoud, Phys. Rev. D **100**, 083528 (2019).
[53] K. Jedamzik, arXiv:2007.03565.
[54] S. Young and A. S. Hamers, J. Cosmol. Astropart. Phys. 10 (2020) 036.
[55] M. Trashorras, J. García-Bellido, and S. Nesseris, arXiv:2006.15018.
[56] L. Liu, Z.-K. Guo, and R.-G. Cai, Eur. Phys. J. C **79**, 717 (2019).
[57] J. Adamek, C. T. Byrnes, M. Gosenca, and S. Hotchkiss, Phys. Rev. D **100**, 023506 (2019).
[58] V. Bosch-Ramon and N. Bellomo, Astron. Astrophys. **638**, A132 (2020).
[59] J. L. G. Sobrinho and P. Augusto, Mon. Not. R. Astron. Soc. **496**, 60 (2020).
[60] A. Dolgov and J. Silk, Phys. Rev. D **47**, 4244 (1993).
[61] B. Carr, M. Raidal, T. Tenkanen, V. Vaskonen, and H. Veermäe, Phys. Rev. D **96**, 023514 (2017).
[62] A. D. Gow, C. T. Byrnes, and A. Hall, arXiv:2009.03204.
[63] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaboration), Phys. Rev. Lett. **118**, 221101 (2017).

[64] E. D. Kovetz, I. Cholis, P. C. Breysse, and M. Kamionkowski, Phys. Rev. D **95**, 103010 (2017).
[65] C. Talbot and E. Thrane, Astrophys. J. **856**, 173 (2018).
[66] F. Özel, D. Psaltis, R. Narayan, and J. E. McClintock, Astrophys. J. **725**, 1918 (2010).
[67] C. Talbot and E. Thrane, Phys. Rev. D **96**, 023012 (2017).
[68] J. Garcia-Bellido, J. F. Nuño Siles, and E. Ruiz Morales, arXiv:2010.13811.
[69] C. Kimball, C. Talbot, C. P. L. Berry, M. Carney, M. Zevin, E. Thrane, and V. Kalogera, Astrophys. J. **900**, 177 (2020).
[70] T. J. Loredo, AIP Conf. Proc. **735**, 195 (2004).
[71] I. Mandel, W. M. Farr, and J. R. Gair, Mon. Not. R. Astron. Soc. **486**, 1086 (2019).
[72] K. Vattis, I. S. Goldstein, and S. M. Koushiappas, Phys. Rev. D **102**, 061301 (2020).
[73] Y.-D. Tsai, A. Palmese, S. Profumo, and T. Jeltema, arXiv:2007.03686.
[74] S. Vitale, R. Lynch, R. Sturani, and P. Graff, Classical Quantum Gravity **34**, 03LT01 (2017).
[75] J. Skilling, in *American Institute of Physics Conference Series* (AIP Publishing, College Park, Maryland, 2004), Vol. 735, pp. 395–405.
[76] J. Skilling, Bayesian Anal. **1**, 833 (2006).
[77] F. Feroz, M. P. Hobson, and M. Bridges, Mon. Not. R. Astron. Soc. **398**, 1601 (2009).
[78] J. S. Speagle, Mon. Not. R. Astron. Soc. **493**, 3132 (2020).
[79] R. Trotta, Mon. Not. R. Astron. Soc. **378**, 72 (2007).
[80] D. J. C. MacKay, *Information Theory, Inference & Learning Algorithms* (Cambridge University Press, Cambridge, England, 2002).
[81] A. Gelman, X. li Meng, and H. Stern, *Statistica Sinica* (Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, 1996), pp. 733–807.
[82] M. Fishbach, W. M. Farr, and D. E. Holz, Astrophys. J. Lett. **891**, L31 (2020),
[83] B. Carr and F. Kühnel, Phys. Rev. D **99**, 103535 (2019).
[84] S. Clesse and J. García-Bellido, arXiv:2007.06481.
[85] G. D. Quinlan and S. L. Shapiro, Astrophys. J. **343**, 725 (1989).
[86] H. Mouri and Y. Taniguchi, Astrophys. J. Lett. **566**, L17 (2002).
[87] M. Ricotti, Astrophys. J. **662**, 53 (2007).
[88] R. Abbott *et al.* (LIGO Scientific and Virgo Collaboration), Phys. Rev. Lett. **125**, 101102 (2020).
[89] S. Bhagwat, V. De Luca, G. Franciolini, P. Pani, and A. Riotto, arXiv:2008.12320.
[90] V. De Luca, V. Desjacques, G. Franciolini, P. Pani, and A. Riotto, arXiv:2009.01728.
[91] T. Akutsu *et al.* (KAGRA Collaboration), arXiv:2008.02921.
[92] LIGO Scientific Collaboration, LIGO Algorithm Library—LALSuite, free software (GPL), 2018.
[93] Wolfram, HypergeometricPFQ, 2001, https://functions.wolfram.com/HypergeometricFunctions/Hypergeometric1F2/06/02/03/0001/.