Using the Binary Black Hole Population to Study Cosmology and the Stochastic Gravitational-Wave Background

Jake Summers
*School of Earth and Space Exploration, Arizona State University, Tempe, AZ 85287, USA**

Mentors: Jacob Golomb and Alan Weinstein
Department of Physics, California Institute of Technology, Pasadena, CA 91125, USA and
LIGO Laboratory, California Institute of Technology, Pasadena, CA 91125, USA

Final Report for the 2024 LIGO SURF Program — v20240925
(Dated: September 25, 2024)

The stochastic gravitational-wave background is the overall signal of gravitational waves produced by numerous astrophysical and cosmological sources, but, in particular, includes the superposition of all binary black hole coalescences in the universe. We use results from LIGO-Virgo-KAGRA to analyze the population of binary black hole coalescences and the gravitational-wave background expected to be caused by them. In analyzing the population of binary black hole coalescences, we consider several mass and redshift population models, using hierarchical Bayesian inference to infer the population. Using these population models, we compute the expected stochastic gravitational-wave background, finding agreement with previous analyses. However, since the catalogs are limited to low-redshift binary black holes, there is significant uncertainty in the empirically-predicted gravitational-wave background. Using a phenomenological model for the merger rate as a function of redshift, we explore how changing the model behavior at high redshifts affects the predicted gravitational-wave background spectrum.

This work uses data from LVK observing runs 1–3, but all work is by the authors and does not represent the LVK collaboration as a whole.

I. INTRODUCTION

A. Compact Binary Coalescences and the Gravitational-Wave Background

In 2015 the Laser Interferometer Gravitational-Wave Observatory (LIGO) made its first detection of a binary black hole merger through the detection of gravitational waves [1]. Since then, the LIGO-Virgo-KAGRA (LVK) collaboration has accumulated a catalog of some $\gtrsim$100 compact binary coalescences (CBCs) [2]. This catalog of CBCs includes many binary black hole (BBH) events, along with a few binary neutron star (BNS) and neutron star–black hole (NSBH) events. With such a large collection of CBCs to analyze, it is possible—and extremely useful—to analyze the population of CBCs as a whole, to obtain deeper insights regarding the set of gravitational wave signals that the LVK detects. In turn, we can learn more about the astrophysical processes that yield these signals.

In particular, population analyses can be used to measure the (observed) mass distribution, spin distribution, redshift distribution, and merger rate of CBCs [2]. Furthermore, these population properties can be used to predict the *overall* population (both observed and unobserved) of CBCs. This population produces the signals that form components of the stochastic gravitational-wave background (GWB). This GWB manifests as a continuous source of gravitational radiation; this is in contrast to CBCs, which are discrete events which are modeled through numerical relativity and post-Newtonian waveform approximates. While not yet observed, the GWB produced by stellar-mass BBH mergers could provide invaluable information about the mostly-unobserved astrophysical population of stellar-mass black holes in our universe.

There are numerous contributors to the overall GWB. Among those, ground-based (*e.g.,* LVK, 20 Hz $\lesssim f \lesssim$ 10^4 Hz) and space-based gravitational-wave detectors (10^{-5} Hz $\lesssim f \lesssim$ 1 Hz) may eventually be able to see a stochastic background consisting of a superposition of cosmological sources (*e.g.,* vacuum fluctuation amplification, cosmic strings, and phase transitions) and astrophysical sources (*e.g.,* triaxial emission and instabilities due to rotating neutron stars, core-collapse supernovae, supermassive black hole capture, and of course—CBCs) [3]. Additionally, pulsar timing data has provided strong evidence for a GWB in the nanohertz frequency range, believed to be caused by the inspiral of supermassive black hole binaries [4].

Using most CBC signals detected by the LVK network and collected in the Gravitational Wave Transient Catalog 3 (GWTC-3), previous analyses have predicted the GWB from BBH, BNS, and NSBH coalescences, under a handful of simplifying assumptions (*e.g.,* Fig. 5 of Ref. [5] and Fig. 23 of Ref. [2]). In particular, these analyses assume that the GWB is isotropic. Given these predictions, it is possible that the GWB from CBCs can

* jssumme1@asu.edu

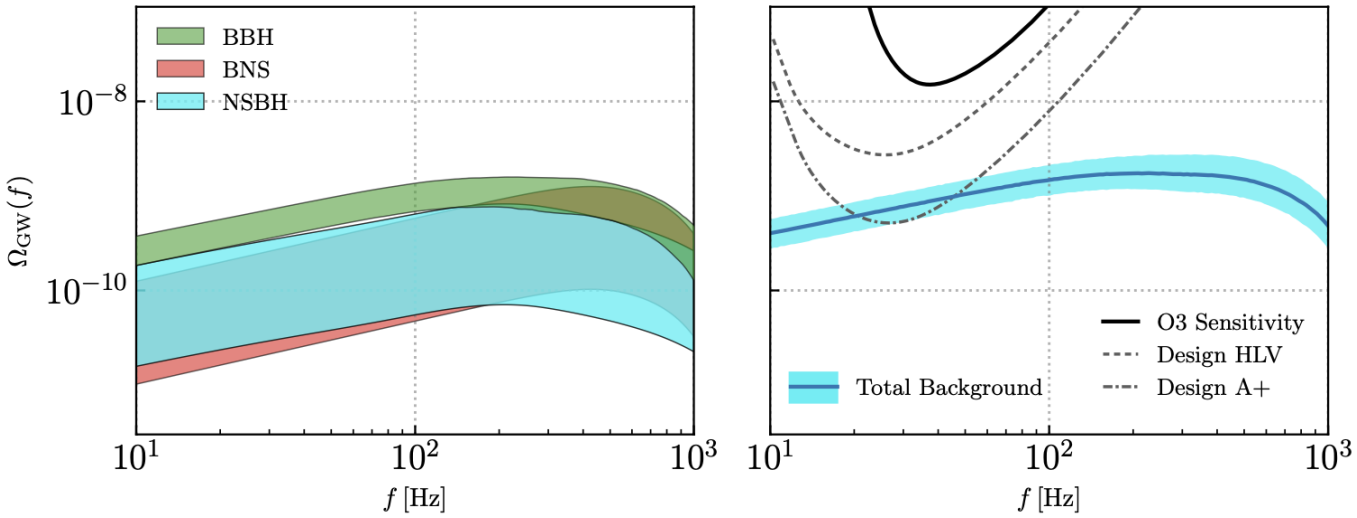

FIG. 1. Prediction of the GWB from CBCs following LVK O3, reproduced from Fig. 23 of Ref. [2]. The GWB is quantified using $\Omega_{\rm GW}(f)$, the dimensionless energy density spectra, and is plotted as a function of frequency (f). The individual contributions from BBH, BNS, and NSBH mergers are plotted on the left; the GWB due to BNS and NSBH mergers is determined by Poisson uncertainty, while the GWB from BBH mergers additionally accounts for uncertainty in the mass distribution. The total GWB estimate is plotted on the right, with additional curves depicting the current LVK sensitivity and the sensitivity of future gravitational-wave detectors under an isotropic GWB search. "Design HLV" stands for Design Hanford-Livingston-Virgo and refers to the design sensitivity of LVK O4; "Design A+" corresponds to the O5 design sensitivity.

be observed at around 30 Hz using the future "A+" LIGO design. Fig. 1 depicts the GWB predictions from Ref. [2], as well as the LVK's current and future sensitivities under an isotropic search. These analyses have used the population of observed BBH coalescences to inform the predicted GWB; however, upper limits on the actual stochastic GWB can also be used to inform population models [6].

In addition to the isotropic GWB, there also exists an anisotropic GWB that consists of well-localized sources that emit GWs. In particular, low-redshift anisotropies, such as galaxy clusters, contribute to the anisotropic GWB. At large scales, the universe is essentially homogeneous, so the GWB becomes isotropic at such scales. However, at smaller scales, the universe is not homogeneous. This inhomogeneity could be used to localize GWs, providing us with complementary info about the GWB. Anisotropic searches are not as sensitive to quiet signals, but do not suffer from an overlap reduction function, so they are sensitive to a wider range of frequencies compared to isotropic searches. Such anisotropic GWB searches have been previously undertaken, but have so far been unsuccessful [7]. Yet, a nondetection of the anisotropic GWB still sets upper-limits on the strength of the signals.

B. LVK Observing Run 4 and Beyond

The LVK collaboration underwent their first period of observing run 4 (O4a) from May 24th, 2023 to January 16th, 2024. Their second observing period, O4b, started on April 10th, 2024, and will run for around nine months. This means that there are many more CBCs detected in O4a and O4b that can soon be used to improve predictions of the GWB from CBCs.

Beyond LVK, plans have been made to build more terrestrial gravitational wave detectors—the Einstein Telescope (ET) and the Cosmic Explorer (CE)—and to begin building space-based detectors—the Laser Interferometer Space Antenna (LISA) [8]. Whereas the LVK collaboration can reach frequencies as low as ∼20 Hz, the ET and CE will be sensitive down to ∼3 Hz. Such improvements will allow for the measurement of coalescences of more massive objects. In fact, while LVK can detect mergers between BBHs with masses of $m \lesssim 200\ M_\odot$, ET and CE could detect masses as large as $m \lesssim 1000\ M_\odot$. Additionally, ET and CE could detect BBHs out to $z \lesssim 20$ (if such events exist), compared to LVK at $z \lesssim 1$ [8].

With these drastic improvements in detection ability, it is possible that such next-generation gravitational wave detectors could detect coalescences between BH remnants from Population III stars (to be discussed shortly) at $z \approx 5 - 20$ with masses $m \approx 100 - 1000\ M_\odot$. Additionally, it is certain that these detectors could detect hierarchical black hole coalescences, which can produce intermediate mass BHs via coalescences between lighter BHs. Other key science questions to be addressed by CE include observation of the complete population of stellar-mass BBHs, characterization of the $\sim 60\ M_\odot - 150\ M_\odot$ mass gap caused by pair instability, investigation of the interiors of neutron stars, and the search for new physics [9].

C. High-Redshift Behavior

Because LVK has only observed BBH coalescences out to $z \lesssim 1$, we are justified in taking other routes in order to *estimate* the high-redshift behavior of BBH coalescences. One such route is to stitch together both electromagnetic observations and simulations to create a hybrid model. Our goal here is not to propose a model that is necessarily correct, but to create a model which is flexible to including unexpected high-redshift behavior, and is in the realm of possibility.

Our first clues come from the *James Webb Space Telescope* (*JWST*). Since *JWST* started returning data, there have been around 20 spectroscopically-confirmed galaxies at $z > 10$ that can tell us about the epoch of Reionization. In particular, these galaxies exhibit a larger ultraviolet luminosity density than expected from previous (e.g., *Hubble Space Telescope*) surveys. Many *JWST* observations have revealed young, massive star clusters in high-redshift galaxies, and in many cases $\gtrsim$10% of these galaxies' total mass is from star clusters. These potential proto-globular clusters could allow for runaway stellar and BH mergers in their dense cores [10].

High-redshift deviations from the standard Madau & Dickinson star-formation rate (SFR) [11] have been previously reported [12]. Such observations using the Subaru/Hyper Suprime-Cam survey and CFHT Large Area U-band Survey have shown that the Madau & Dickinson SFR *overpredicts* the observed SFR density for $z \gtrsim 7$. Yet, more recent observations done with *JWST* have seen a SFR density that overpredicts the model derived from Subaru and CFHT data for $z \gtrsim 10$ [13]. Thus the *JWST* survey could point towards an overabundance of BBH coalescences at mid-high redshifts (the time-delay between star formation and black hole merger *could* allow high-redshift star-formation to transfer over to a higher observed merger rate at intermediate redshifts).

An additional piece in understanding the potential high-redshift behavior of BBH coalescences is obtained through simulations. We previously mentioned that high-redshift galaxies exhibit an overabundance of massive proto-globular clusters. Then, it makes sense to look into simulations of globular clusters (GCs) to understand how their dynamical BBH coalescences are different than the typical field-formation scenario. In terms of the mass spectrum produced by BBH coalescences in GCs, Ref. [14] finds that such coalescences can reproduce the Power Law + Peak (PP) model (see Sec. IV or App. A), *except for the low-mass peak*. They also find that high-density GCs ($\sim 10^5\ M_\odot\ \mathrm{pc}^{-3}$) do the best job at reproducing the PP model, and that hierarchical mergers are responsible for all coalescences with $m_1 \gtrsim 50\ M_\odot$.

While current LVK sensitivities do now allow for us to observe the high-redshift behavior that we have discussed in this section, future LVK observing runs and future detectors will observe many more BBH coalescences, allowing us to directly extract information about the high-redshift SFR and the contribution of GC dynamical interactions towards the BBH merger rate and mass spectrum.

D. Population III Stars

Our universe is populated by three primary categories of stars, which happen to be named in backwards order. Population I stars are those like our sun that have solar metallicities and exist in the disks of galaxies. Population II stars are older stars which have lower metallicities and exist around galactic bulges and halos. The Population II stars that underwent supernovae seeded the material which went into Population I stars. Finally, Population III (hereafter Pop. III) stars are thought to be stars with essentially zero metallicity—those made entirely from primordial Hydrogen gas. Simulations provide limited information about the potential properties of Pop. III stars, and we have yet to observe any Pop. III stars directly with electromagnetic observations. Hence, it is conceivable that GW observations can help constrain the properties of Pop. III stars through the GWB.

For now, we can only assume an astrophysically-motivated initial mass function (IMF) for Pop. III stars and use that to predict how these stars can influence the GWB. However, the IMF of Pop. III stars is ill-constrained. Simulations predict a relatively (logarithmically) flat IMF ranging from 0.01 $M_\odot$ up to 1000 $M_\odot$; yet, models that include radiative feedback predict higher masses than those that don't [15]. The IMF depends on the poorly-understood fragmentation of primordial accretion disks. Simulations that model a longer period of time observe less fragmentation due to their inherently lower resolution, so they are biased towards predicting larger stars [15]. The fragmentation of primordial gas also causes many Pop. III stars to form in multiple-star systems, which could increase the chance of a BBH coalescence to occur.

At the end of their lives, Pop. III stars of adequate mass will form BHs. Pop. III stars with $25\ M_\odot < M < 40\ M_\odot$ will undergo a core-collapse supernovae; however, such a supernova would be faint because the stellar envelope does not completely unbind from the star (as opposed to less-massive Pop. III stars, which go supernovae and leave behind a NS). Larger Pop. III stars below $\sim$100 $M_\odot$ can directly collapse into a BH. Stars larger than $\sim$100 $M_\odot$ (this number is quite uncertain), but smaller than 140 $M_\odot$ will undergo pulsational pair instability, where highly-energetic photons in the core convert into electron-positron pairs, creating a runaway nuclear reaction that causes the core to implode; then, the star releases a flash of energy, allowing the star to reexpand. The process of pulsational pair instability causes the star to release energy until its mass is small enough to undergo the usual evolution sequence. Between 140 $M_\odot$ and 260 $M_\odot$, the pair instability is strong enough to blow up the entire star in a pair instability supernovae, so no BH is left behind. Finally, above 260 $M_\odot$, the star is able

to collapse directly into a BH, despite pair instability still being present [15]. Note that the processes discussed here are not necessarily unique to Pop. III stars and can also occur for Pop. I and Pop. II stars, albeit at different mass ranges

Despite their high-redshifts, it is conceivable that Pop. III stars could eventually be detected by next-generation GW detectors. In the meantime, we can try to constrain their characteristics with LVK.

This paper is organized as follows. In Sec. II, we discuss the details behind BBH population inference and the stochastic GWB. In Sec. III, we discuss the data used. In Sec. IV, we discuss the codes used to perform population inference and calculation of the GWB, as well as the population models used to model the BBH population. In Sec. V, we discuss our results from population inference. In Sec. VI, we discuss our findings regarding the stochastic gravitational-wave background generated by our catalog of BBH coalescences. In Sec. VII, we take a phenomenological approach to analyzing the GWB. Finally, in Sec. VIII, we conclude and discuss opportunities for the future.

We elaborate on the mass, spin, redshift, and evolving models in App. A, App. B, App. C, and App. D, respectively.

In this paper, we use standard Planck cosmology [16] with $H_0 = (67.4 \pm 0.5)$ km s^{-1} Mpc^{-1} and $\Omega_m = 0.315 \pm 0.007$.

Unless otherwise indicated in the figure caption, all plots are made by the authors. The code used to produce the figures can be found on GitLab.[1]

II. BACKGROUND

A. Population Inference

The goal of population inference is to search for valid models from which all events that occur (accounting for observational bias) are drawn from. Such models can be used to tell us about the astrophysical properties of the population of BBHs we observe, and also better-inform our parameter estimation (once we model the population, those models can then be used to update our priors). To model the population, it is standard in gravitational-wave astronomy to use hierarchical Bayesian inference. This framework defines the likelihood of an individual BBH event, given the hyper-parameters Λ of the population model, as [17]

$$\mathcal{L}(d_i|\Lambda) = \int \mathrm{d}\theta \; \mathcal{L}(d_i|\theta)\pi(\theta|\Lambda), \tag{1}$$

where, in practice, we do not directly know $\mathcal{L}(d_i|\theta)$ from the catalog. However, from parameter estimation of individual BBH events, we can assume that it can be retrieved from the posterior distribution and the prior. With this technique, along with Monte Carlo integration, the likelihood for an individual event can be approximated as [17]

$$\begin{aligned}\mathcal{L}(d_i|\Lambda) &= \int \mathrm{d}\theta \; \mathcal{Z}(d_i)\frac{p(\theta|d_i)}{\pi(\theta)}\pi(\theta|\Lambda) \\ &\approx \frac{\mathcal{Z}(d_i)}{m}\sum_{\theta_j \sim p(\theta|d_i)}^{m}\frac{\pi(\theta_j|\Lambda)}{\pi(\theta_j)},\end{aligned} \tag{2}$$

where $\mathcal{Z}(d_i)$ is the evidence calculated during parameter estimation for the event. The integral is approximated by summing over m events drawn from the prior distribution $p(\theta|d_i)$. $\pi(\theta_j)$ is a default prior used for parameter estimation that is saved in the catalog; as long as the prior is sufficiently wide, no erasure of information will take place. When we have more than one event in the population (we use 71 events in GWTC-3), the total likelihood is calculated as the product of all of the individual likelihoods, divided by a selection effects term, and is marginalized over the rate with a $1/\mathcal{R}$ prior [18]:

$$\begin{aligned}\mathcal{L}(\{d_i\}|\Lambda) &\propto \prod_i^N \frac{\mathcal{L}(d_i|\Lambda)}{P_{\mathrm{det}}(\Lambda)} \\ &\propto P_{\mathrm{det}}(\Lambda)^{-N}\prod_i^N \mathcal{L}(d_i|\Lambda).\end{aligned} \tag{3}$$

The selection effects term, $P_{\mathrm{det}}(\Lambda)$, is the fraction of all signals that are detectable from a population with hyper-parameters Λ (*i.e.,* Malmquist bias). This term is calculated by injecting signals into the LVK data, and saving the parameters of the signals that are detected (using the same detection threshold and the same data and noise distribution as with real events). Numerically, it is

$$P_{\mathrm{det}}(\Lambda) = \int \mathrm{d}d\mathrm{d}\theta \; \mathcal{L}(d|\theta)\pi(\theta|\Lambda)p_{\mathrm{det}}(d), \tag{4}$$

where $\mathcal{L}(d|\theta)$ is the likelihood, $\pi(\theta|\Lambda)$ is the prior, and $p_{\mathrm{det}}(d)$ is specific to each injected event, so it is either zero or one. This can be computed by once again using Monte Carlo integration:

$$P_{\mathrm{det}}(\Lambda) \approx \frac{1}{N_{\mathrm{inj}}}\sum_{\theta \sim \theta_{\mathrm{found}}}\frac{p(\theta|\Lambda)}{\pi(\theta|\Lambda_0)}, \tag{5}$$

assuming there are N_{inj} total injections, and of those, we draw parameters θ_{found} only from the events that were detected ($p_{\mathrm{det}}(d) = 1$). In this case $\pi(\theta|\Lambda_0)$ is a fiducial prior used for reweighting.

Once the likelihood is calculated, and a prior on Λ is decided upon (usually uniform), an appropriate stochastic sampler can be chosen to compute the posterior distribution $p(d|\Lambda)$. The two major options for a sampler are Markov-chain Monte Carlo (MCMC) and nested sampling. With MCMC, particles undergo a random walk

[1] `https://git.ligo.org/jake.summers/astro-gwb`

through the posterior to generate samples. Since the samples are strongly dependent on the initial conditions of the walkers, there is usually a "burn-in" phase to allow the walker to find a steady state before actually generating samples from the posterior. With nested sampling, some number of "live points" are taken from the prior, and the lowest likelihood points are gradually removed in favor of high-likelihood points. Nested sampling techniques are able to estimate the evidence directly, unlike basic MCMC techniques [17]. In the end, we obtain a finite set of samples that reflect the posterior.

B. The BBH Gravitational-Wave Background

The stellar-mass BBH stochastic GWB is the overall signal of gravitational radiation, from a superposition of all stellar-mass BBH coalescences in the universe. LVK can only detect the loudest BBH events, but all undetectable events would contribute stochastic—but correlated—signal to the LVK detectors. This stochastic GWB is currently undetected by the LVK, but it can be estimated through population inference. The fractional energy density spectrum emitted by gravitational waves is characterized by [2]

$$\Omega(f) = \frac{1}{\rho_c}\frac{\mathrm{d}\rho}{\mathrm{d}\ln f}, \tag{6}$$

where $\rho_c = 3c^2H_0^2/8\pi G$ is the critical energy density to make the universe flat. Intuitively, $\Omega(f)$ can be thought of as the cosmological density parameter for gravitational waves. Since $\Omega(f)$ is a unitless fraction of total energy density, it is useful to write down the energy spectrum associated with an individual gravitational-wave event [19]:

$$\frac{\mathrm{d}E}{\mathrm{d}f} = f^2\frac{\pi c^3}{2G}\int \mathrm{d}A\ \left[\left|\tilde{h}_+(f)\right|^2 + \left|\tilde{h}_\times(f)\right|^2\right], \tag{7}$$

where $\tilde{h}_+$ and $\tilde{h}_\times$ are the Fourier transforms of the "+" and "×" polarizations of the metric perturbation, respectively. The integral is over the surface of the unit sphere surrounding the source of the gravitational waves. In the detector frame, $\Omega(f)$ can be estimated as the average gravitational-wave energy present; in practice, this estimation must have an associated observation time $T_{\rm obs}$ [19]:

$$\Omega_{\rm GW}(f) = \frac{f^3}{T_{\rm obs}}\frac{4\pi^2}{3H_0^2}\sum_i^{N_i}\mathcal{P}_d(\Theta_i; f). \tag{8}$$

In this equation, $\mathcal{P}_d(\Theta_i; f)$ is the Fourier domain unpolarized power in the detector frame for a gravitational wave with parameters Θ_i, $T_{\rm obs}$ is the observing time, and N_i is the number of events. This signal is inherently time-dependent, and will change for different $T_{\rm obs}$ and for different reference times. However, given infinite time, $\Omega_{\rm GW}$ will approach its ensemble average, $\bar{\Omega}_{\rm GW}$. The ensemble-averaged gravitational-wave energy density is a population-dependent quantity. In the limit of infinite N_i and infinite $T_{\rm obs}$, the ensemble average is [19]

$$\bar{\Omega}_{\rm GW}(\Lambda; f) = f^3\frac{4\pi^2}{3H_0^2}\mathcal{R}\int \mathrm{d}\Theta\ p_d(\Theta|\Lambda)\mathcal{P}_d(\Theta; f), \tag{9}$$

where $\mathcal{R} = \mathrm{d}N/\mathrm{d}t$ is the merger rate in the detector frame and $p_d(\Theta|\Lambda)$ is the detector frame probability distribution of the CBC parameters Θ informed by the hyper-parameters Λ. In practice, this can be calculated by computing energies for a grid of waveforms, along with their model-dependent probability distributions, and performing Monte Carlo integration. However, calculating $\bar{\Omega}_{\rm GW}$ can be very computationally intensive when one has redo the calculation on multiple sets of hyper-parameters. Thus, Ref. [19] employs a reweighting technique such that

$$\begin{aligned}\int \mathrm{d}\Theta\ p_d(\Theta|\Lambda_1)\mathcal{P}_d(\Theta) &= \int \mathrm{d}\Theta\ \frac{p_d(\Theta|\Lambda_1)}{p_d(\Theta|\Lambda_0)}p_d(\Theta|\Lambda_0)\mathcal{P}_d(\Theta)\\ &\equiv \int \mathrm{d}\Theta\ w_{0\to1}(\Theta)p_d(\Theta|\Lambda_0)\mathcal{P}_d(\Theta)\\ &\approx \sum_j w_{0\to1}(\Theta_j)\mathcal{P}_d(\Theta_j),\end{aligned} \tag{10}$$

where $w_{0\to1}$ is the weight, which can be used to quickly recalculate the leftmost integral. When Monte Carlo integration is used here for sufficiently large j, the waveform power $\mathcal{P}$ only needs to be evaluated once, then additional calculations can be performed by simply recalculating the weight factors.

III. DATA

In this paper, our primary catalog is the Gravitational Wave Transient Catalog 3 (GWTC-3), which includes posteriors for all signals in O1, O2, and O3 [20]. In GWTC-3, we use 71 BBH coalescences with a false alarm rate (FAR) threshold $< 1\ \mathrm{yr}^{-1}$. This threshold is somewhat generous, but it is reasonable for such a large population of black holes.

In addition to the catalog of BBH coalescences, we also use injection data to characterize the probability of detecting various types of BBH events. For the GWTC-3 catalog, we use a catalog of $\sim 4\times10^4$ detected injections, out of $\sim 7\times10^7$ total injections. Note that these injections are cumulative, and approximately weighted to represent the O1, O2, and O3 sensitivities properly.

It should be noted that the injection data only allows for $M \leq 100\ M_\odot$ black holes, which places a restriction on the population models and events we can use. Because of this, we restrict our analysis to BBH events with $m_1 \leq 100\ M_\odot$. Additionally, we exclude BNS and NSBH events, as well as any events which may throw off the population inference. With enough detections, population inference will be able to help us better-interpret

outlier events, but at the present, the models used are not equipped to identify outliers with confidence.

IV. METHODS

A. Population Inference with `gwpopulation`

To perform population inference, as described in Sec. II, we use the python module `gwpopulation`[2] [21]. With this, we use models native to `gwpopulation`, with sufficiently wide, uniform priors, on all hyper-parameters; the specific models that we use are described below, however see the Appendix for more details. In the likelihood function, we restrict the maximum variance in the log-likelihood to be ≤ 1, because otherwise, narrow features in the population are unreasonable favored (see Ref. [18] for a discussion regarding uncertainty). We confirmed that using a maximum uncertainty of two does not significantly affect the results. We use nested sampling with the `dynesty` sampler, with 500 live points. The actual sampling is performed using `bilby`[3] [22]. We also use `gwpopulation_pipe`[4] [23] to compute the local merger rate posterior after the nested sampling finishes.

B. GWB Computation with `popstock`

To calculate the ensemble GWB from stellar-mass BBH coalescences using the reweighting technique, we use `popstock`[5] [19]. For each GWB computation, we use a representative sample of 5×10^5 BBH coalescences taken from the population; the primary reason for this number is that too few samples can result in bias due to small number statistics, and too many samples is too computationally intensive. We save these samples for each specific combination of models to ensure consistency between runs. We also save the corresponding wave energies (a computation that must take place on the CPU) to allow us to later perform reweighting on the GPU with `jax` [24]. With this, we incorporate uncertainty in the GWB from all available mass and redshift model parameters. We do not include uncertainty from the spin model, as the GWB is not very sensitive to the spin model. Finally, we include uncertainty in the local merger rate. Note that the GWB credible intervals are often limited by the prior because high-redshift hyper-parameters are not well-constrained. To calculate the overall GWB from the BBH model, we take the posteriors from population inference, calculate $\bar{\Omega}_{\rm GW}$ for each posterior draw (using reweighting), and then generate confidence intervals from those. This results in an efficient way to compute the GWB.

C. Mass Models

Here, we qualitatively describe the mass models used in this paper. For a complete description of these models, see App. A. There are three mass models that we use, which we describe in turn.

The Power Law + Peak (hereafter PP) model describes the primary mass (m_1) distribution as a superposition between a power-law distribution and a Gaussian peak; it describes the mass ratio (q) with a simple power-law [25]. Both the primary mass and mass ratio distribution are smoothed at low masses. The model hyper-parameters describe the spectral index of the primary mass power-law (α), the spectral index of the mass ratio power-law (β), the minimum ($m_{\rm min}$) and maximum ($m_{\rm max}$) mass, the fraction of primary masses contained in the Gaussian peak (λ_m), the mean (μ_m) and standard deviation (σ_m) of the peak, and the width of the smoothing function (δ_m). The peak in the PP model is physically motivated by pulsational pair-instability supernovae (PPISNe) [25], but is probably caused by something else [26] because the mean mass of the peak is too low to be consistent with being caused by PPISNe.

The Primary-Secondary Identical (hereafter M1M2) model also uses a power-law with a Gaussian peak, but instead of modeling the primary mass and mass ratio, it models the primary mass and secondary mass. It models these two distributions based on the single power-law + peak distribution (which describes the *underlying* population of BHs). This model uses the same hyper-parameters as the PP distribution, except for the fact that it lacks the β parameter to describe a power-law in mass ratio. This model is used to demonstrate that a pairing function is necessary to describe how BHs pair up with each other [27].

The Paired Primary-Secondary (Paired M1M2) model attempts to fix the pitfalls of the M1M2 model by introducing a pairing function. The joint probability distribution between the primary and secondary masses is now in the form of $p(m_1, m_2|\Lambda) \propto p(m_1)p(m_2)f(q, M)$, where the pairing function f is a function of the mass ratio (q) and total mass (M). We take our pairing function to be $f(q, M|\Lambda) = q^{\beta_q} M^{\beta_M}$ (as in Ref. [27]), but set a narrow prior on β_M centered around zero because β_M is degenerate with the mass model power-law index α.

Our implementations of the M1M2 and Paired M1M2 models are publicly available on GitHub[6].

[2] `https://github.com/ColmTalbot/gwpopulation`
[3] `https://git.ligo.org/lscsoft/bilby`
[4] `https://git.ligo.org/RatesAndPopulations/gwpopulation_pipe/`
[5] `https://git.ligo.org/arianna.renzini/popstock`
[6] `https://github.com/jssumme1/gwpopulation/blob/m1m2models/gwpopulation/models/mass.py`

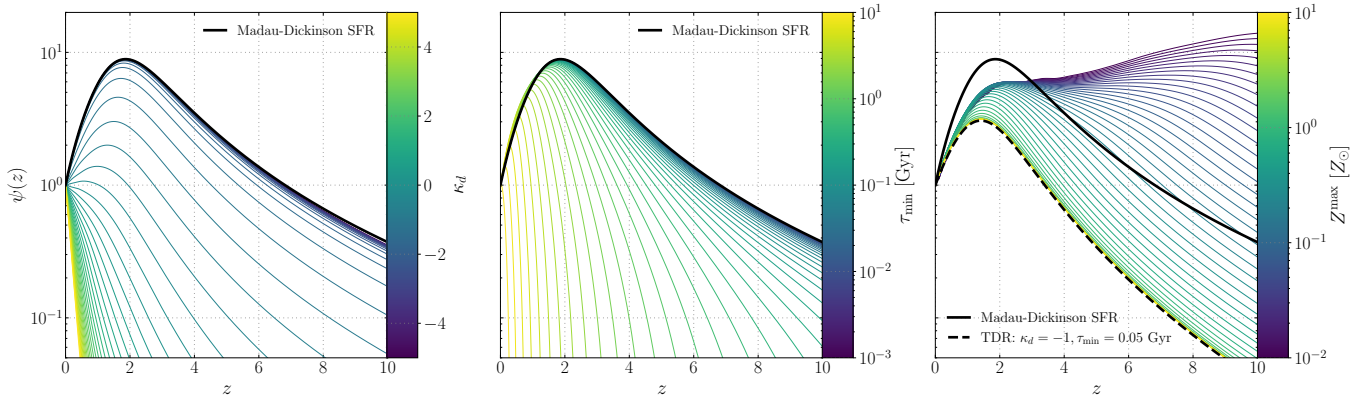

FIG. 2. Normalized merger rate (ψ) as a function of redshift (z) under the Metallicity-Weighted Time Delay Redshift model. The leftmost plot shows curves of varying κ_d for a time delay probability distribution of $p(\tau) = \tau^{\kappa_d}$, with a fixed minimum time delay of $\tau_{\rm min} = 0.01$ Gyr. The middle plot shows curves of varying $\tau_{\rm min}$ with fixed $\kappa_d = -100$ (essentially $\tau = \tau_{\rm min}$). Fixing κ_d and $\tau_{\rm min}$ to these values allows the TDR curves (on one end of the spectrum) to approach the Madau-Dickinson star-formation rate, which is plotted in black. The rightmost plot shows curves of varying $Z^{\rm max}$, with κ_d and $\tau_{\rm min}$ fixed to their typical values. Note that the left and middle plots do not incorporate metallicity effects—akin to the limit of $Z^{\rm max} \to \infty$.

D. Spin Models

As with the mass models, we will qualitatively describe the spin model here. See App. B for more details.

The Independent Identical Distribution model (hereafter IID) assumes that spin magnitudes are drawn from a Beta distribution (identical for each BH), and that the spin angles are drawn from a mixture of an isotropic component and an aligned component (called the "Default spin model" in Ref. [2]). More specifically, the dimensionless spin magnitudes a_1 and a_2 are drawn from a Beta distribution determined by the hyperparameters α_χ and β_χ. The spin angles $\cos\theta_1$ and $\cos\theta_2$ are drawn from a Gaussian of magnitude proportional to ξ, centered around $\cos\theta = 1$, with standard deviation σ, summed together with a flat component of magnitude proportional to $(1-\xi)$. The GWB is not very sensitive to the spin distribution (plus spins are not well-constrained in individual events), so we use the IID model for all computations in this paper—exploring how changing the spin model affects the GWB is beyond the scope of this paper.

E. Redshift Models

When the mass distribution and redshift distribution are independent, the joint probability distribution between mass and redshift satisfies [28]

$$\frac{{\rm d}N}{{\rm d}m_1 {\rm d}m_2 {\rm d}z} \equiv \mathcal{R} p(m_1, m_2, z) = \mathcal{R} p(m_1, m_2) p(z). \quad (11)$$

Then the probability density of mergers for a given redshift model $\psi(z)$ is given by [2, 21]

$$p(z|\Lambda) = \frac{\mathcal{C}}{1+z} \frac{{\rm d}V_c}{{\rm d}z} \psi(z|\Lambda), \quad (12)$$

where $\mathcal{C}$ is a normalization factor obtained through numerical integration and ${\rm d}V_c/{\rm d}z$ is the differential comoving volume, which is specific to the adapted cosmological model. Note that the merger rate can be directly calculated from the local merger rate and ψ as $\mathcal{R}(z) = \mathcal{R}_0 \psi(z)$.

Below, we briefly discuss the redshift models used in this paper. See App. C for the full details.

The Power Law Redshift (hereafter PLR) model is the most simple redshift model, parameterizing the redshift distribution as a power-law with only one parameter λ_z. This model works fine for the current low-redshift LVK observations, but is impractical for computing the GWB since it diverges at large redshifts.

The Madau-Dickinson Redshift (hereafter MDR) model uses the same functional form as the star-formation rate [11]. It uses a parameter γ to describe the low-redshift power-law, $z_{\rm peak}$ for the peak redshift, and κ to describe the high-redshift behavior (which goes as a power-law of $\gamma - \kappa$) [28].

The Metallicity-Weighted Time Delay Redshift model (hereafter MWTDR; we abbreviate the model that is not weighted by metallicity as TDR) first assumes the star-formation rate, then says that BBH mergers occur according to some time delay distribution $p(\tau) = \tau^{\kappa_d}$ with $\tau_{\rm min} \leq \tau \leq H_0^{-1}$. This model also assumes a maximum metallicity $Z^{\rm max}$ [29].

In Fig. 2, we illustrate the behavior of the MWTDR model under varying time delay spectral index and minimum time delay. It is clear that in the limit of $\kappa_d \to -\infty$, $\tau_{\rm min} \to 0$, and $Z^{\rm max} \to \infty$, the MWTDR model approaches the SFR. Then increasing minimum time delay,

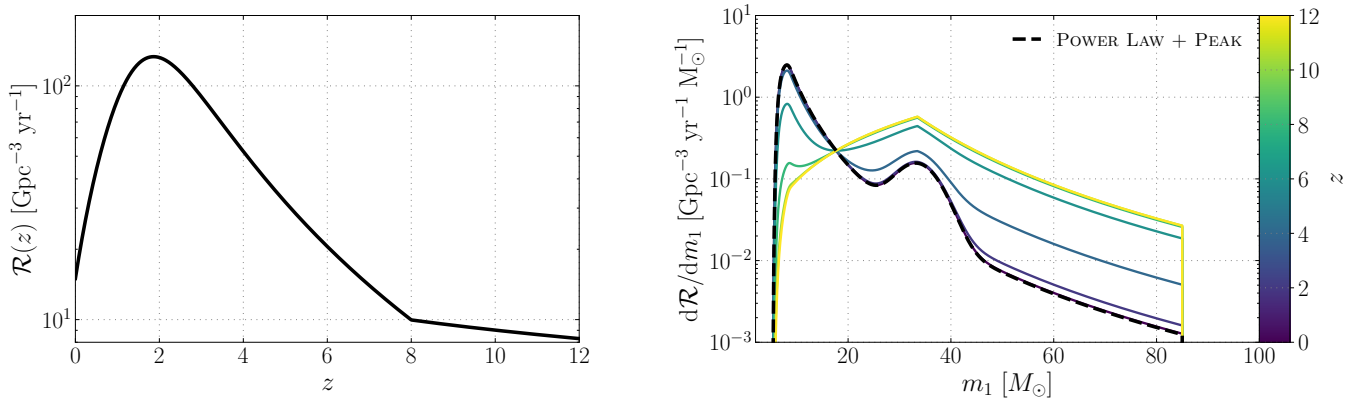

FIG. 3. This figure illustrates an example Power Peak Globular Cluster redshift and mass model. The leftmost plot shows the merger rate ($\mathcal{R}$) as a function of redshift (z) under the PPGC model. The rightmost plot shows the differential merger rate ($d\mathcal{R}/dm_1$) as a function of primary mass (m_1) for varying redshift slices. It is clear that at $z \sim 0$, the mass model reduces to the PP model, and at $z \gtrsim 10$, the model becomes entirely dominated by the GC model.

increasing the spectral index, and decreasing the maximum metallicity causes the MWTDR to further deviate from the SFR.

Our implementation of the MWTDR model is publicly available on GitHub[7].

F. Evolutionary Models

When we go on to analyze the structure of the GWB in a more phenomenological manner, we can suppose that the population of BBHs evolves over time. Such behavior would not be easily detectable with current data, but could affect the GWB when it is eventually detected. We describe our evolutionary models below (see App. D).

The Evolving Power Law + Peak (hereafter EvoPP) model is similar to the PP model, but we also allow some of the model hyper-parameters to linearly change as a function of redshift:

$$\alpha(z) = \alpha_0 + \alpha_z z \tag{13}$$
$$\lambda_m(z) = \lambda_{m,0} + \lambda_{m,z} z \tag{14}$$
$$\mu_m(z) = \mu_{m,0} + \mu_{m,z} z. \tag{15}$$

As a notation simplification, we occasionally denote $\lambda_{m,z}$, for example, by $\lambda(z)$.

The Power Peak Globular Cluster (hereafter PPGC) model is motivated by Sec. I. At low-redshifts, our model should to reduce to the currently-observed population, so it looks like the PP mass model and MDR model at low-redshifts. *JWST* has observed an abundance of massive GCs at high-redshift, which could lead to runaway BBH coalescences [10]. Therefore, supposing that BBH coalescences in the field are subdominant to BBH coalescences from proto-GCs at high-redshift, we can imagine a mass model that is created from the physics in GCs at high-redshifts. We can do this by imposing a logistic sigmoid function that describes the fraction of PP mergers versus GC mergers with central redshift $z_{\rm turn}$ and width $\Delta z_{\rm turn}$. The shape of our GC model is motivated by simulations, which show that the GC mass model looks like the PP model, but without a peak at low-masses [14]. Thus, we describe the GC model as a broken power-law with its break at the same mass as the peak in the PP model. This model has two power-law indices (g_1 and g_2), but in general follows the already-established PP model hyper-parameters.

The redshift component of the PPGC model is motivated by more results from *JWST*. Ref. [13] finds that *JWST* observes more star formation than predicted at high redshifts. Thus, at the same redshift as $z_{\rm turn}$, we impose a break in the MDR redshift model. We do this in a simple manner by just stitching a power-law into the redshift model that starts at this turnover redshift. We add in a power-law of the form $\psi(z \geq z_{\rm turn}) \propto (1+z)^\eta$, and we force this power-law to be continuous with the MDR model at $z = z_{\rm turn}$. However, we additionally implement a boosting factor B that multiplies the merger rate at $z \geq z_{\rm turn}$; this factor can be used to emulate bursty star-formation at very high redshifts.

Our implementations of the EvoPP and PPGC models are publicly available on GitHub[8].

V. POPULATION RESULTS

Before computing the stochastic GWB for a population of BBH coalescences, we first performed population

[7] https://github.com/jssumme1/gwpopulation/blob/m1m2models/gwpopulation/models/redshift.py

[8] https://github.com/jssumme1/gwpopulation/blob/m1m2models/gwpopulation/models/

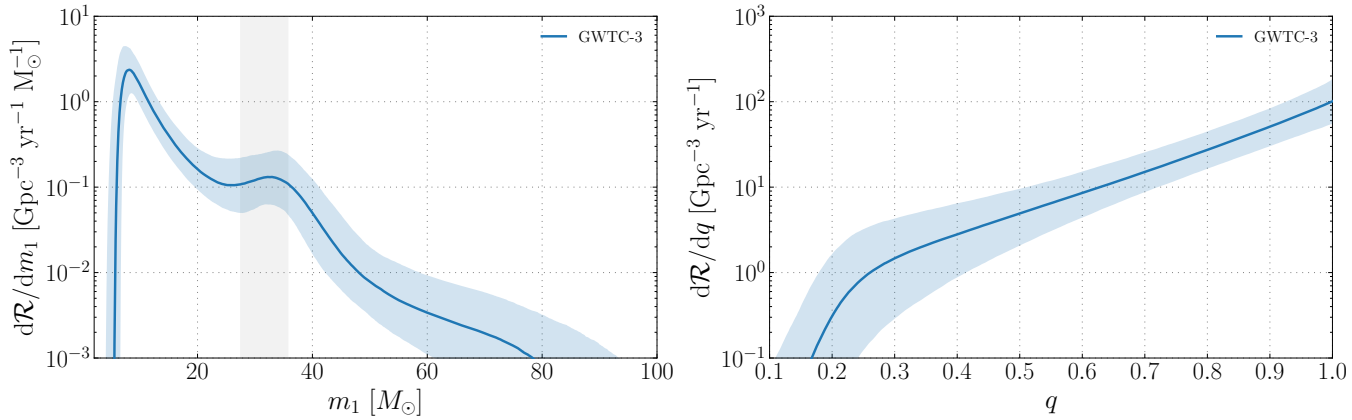

FIG. 4. BBH differential merger rate ($\mathrm{d}\mathcal{R}/\mathrm{d}m_1$ on the left and $\mathrm{d}\mathcal{R}/\mathrm{d}q$ on the right) as a function of primary mass (m_1) and mass ratio (q) for the Power Law + Peak model. The solid blue line shows the median and the shaded regions show the 90% credible intervals for GWTC-3. The shaded grey region on the leftmost plot shows the 90% credible interval for the mean of the Gaussian peak in GWTC-3.

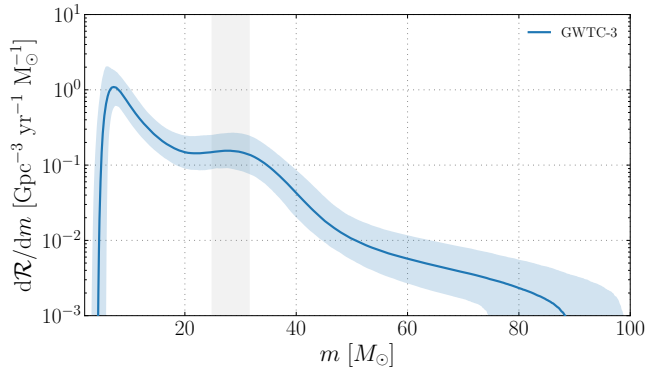

FIG. 5. BBH differential merger rate ($\mathrm{d}\mathcal{R}/\mathrm{d}m$) as a function of mass (for the *underlying* BH mass distribution) for the Primary-Secondary Identical model. The solid blue line shows the median and the shaded region shows the 90% credible intervals for GWTC-3. The shaded grey region shows the 90% credible interval for the mean of the Gaussian peak in GWTC-3.

inference on our catalog. Below we discuss the resultant mass and redshift models; analysis of the spin models is beyond the scope of this paper, but has little affect on the GWB and is weakly constrained by LVK anyways.

A. Mass Distribution

1. PP Model

In Fig. 4, we plot the differential merger rate, $\mathrm{d}\mathcal{R}/\mathrm{d}m_1$. This quantity can be straightforwardly calculated from the local merger rate and the probability distribution for the primary mass. Note that the merger rate is defined as

$$\mathcal{R} = \frac{N}{VT}, \tag{16}$$

where N is the number of events and VT is the product between the volume that LVK is observing and the time spent observing. From this, we can see that

$$\frac{\mathrm{d}\mathcal{R}}{\mathrm{d}m_1} = \frac{1}{VT}\frac{\mathrm{d}N}{\mathrm{d}m_1} = \frac{N}{VT}\frac{1}{N}\frac{\mathrm{d}N}{\mathrm{d}m_1} \equiv \mathcal{R}p(m_1|\Lambda), \tag{17}$$

meaning that the differential merger rate in Fig. 4 is calculated from the product of the local merger rate and the probability distribution. For the mass ratio plot, we have to marginalize over the primary mass since the mass ratio distribution depends on m_1 (the mass ratio has a conditional minimum on m_1).

2. M1M2 Model

Primary-Secondary mass models are not used often in the literature. This is because it is usually necessary to use some sort of pairing function to characterize how BHs obtain a partner. The primary mass–mass ratio models such as the PP model do this easily by assuming the mass ratio goes as a power-law. However, for the M1M2 model, it is more complicated; the model itself is simple, and it is most likely inaccurate because of that. An ideal M1M2 model would have a pairing function that constrains both the mass ratio and the total mass [27]. However, modeling this simple M1M2 model can illustrate the problems of assuming that both BBH components come from the same source.

We do the same population analysis here as with the PP model, and the underlying mass model is plotted in Fig. 5. Given that it uses the same overall functional form as in Fig. 4, the shape is similar. However, there are a few

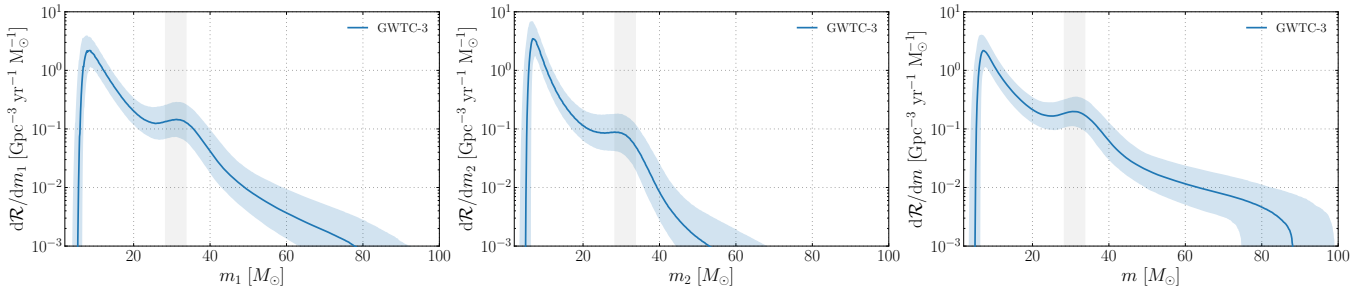

FIG. 6. BBH differential merger rate ($\mathrm{d}\mathcal{R}/\mathrm{d}m$) as a function of mass for the PAIRED PRIMARY-SECONDARY model and for the (left) primary mass, (middle) secondary mass, and (right) underlying mass distribution. The solid blue lines show the medians and the shaded regions show the 90% credible intervals for GWTC-3. The shaded grey regions show the 90% credible interval for the mean of the Gaussian peak for the underlying mass distribution in GWTC-3.

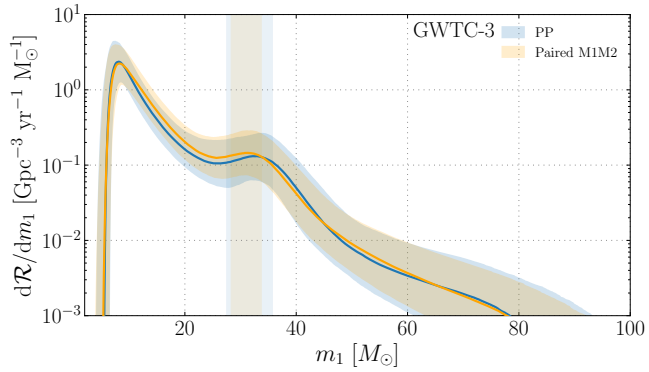

FIG. 7. BBH differential merger rate ($\mathrm{d}\mathcal{R}/\mathrm{d}m_1$) as a function of primary mass (m_1) for the POWER LAW + PEAK model and the PAIRED PRIMARY-SECONDARY model (with GWTC-3). The solid blue and orange lines show the medians for the PP and Paired M1M2 models, respectively. The shaded regions show the 90% credible intervals for the PP and Paired M1M2 models. The vertical shaded regions shows the 90% credible intervals for the mean of the Gaussian peak for each model.

notable differences between this underlying mass model, and the PP primary mass model. The power-law spectral index is around 0.7 lower, the maximum mass is quite a bit higher, the fraction of BHs in the Gaussian peak is much higher, the peak was shifted to lower masses, and the peak is slightly wider. The redshift distribution also changed with the M1M2 model, as the local merger rate is smaller while the low-redshift spectral index is larger. The peak behavior can be explained by both mass distributions having their own peak, and the overall peak just comes out to be the average of the two. Since the PP model does not model the underlying mass distribution ($\mathrm{d}\mathcal{R}/\mathrm{d}m$), this comparison is not entirely one-to-one, but it illustrates the unique information obtained from modeling the underlying mass distribution.

3. *Paired M1M2 Model*

As discussed above, the Paired M1M2 model should be a better model than the M1M2 model because it uses a pairing function. Therefore, we analyze the population of BBH coalescences from GWTC-3 using this model. In Fig. 6, we show the primary mass, the secondary mass, and the underlying mass distributions under this model. The plots for the primary and secondary masses were obtained by taking the joint probability distribution and integrating over the secondary mass and primary mass, respectively. Here, it is noticeable that the primary mass distribution contains higher masses (as it should) than the secondary mass distribution. What is most interesting, however, is that we now have the *underlying* mass distribution in the rightmost panel.

In Fig. 7, we show the PP primary mass model and the Paired M1M2 primary mass model on the same plot for comparison. The two models are generally compatible, showing that our models are sufficiently similar. Yet, the Paired M1M2 model additionally provides information about the underlying distribution, while the PP model does not.

In Fig. 8, we show the two-dimensional merger rate distribution. The benefit of using the Paired M1M2 model is that this plot comes naturally from the population model (since m_1 and m_2 are the two population variables). We can still do the same analysis, however, if we set $q = m_2/m_1$ in the PP model for plotting purposes. In these plots, it is apparent that the Paired M1M2 and PP models have slightly different structures. The Paired M1M2 model has a sloped bottom because of the q_{min} cutoff. It is also interesting how in the Paired M1M2 plot, the peak at $\sim$35 $M_\odot$ looks stronger for m_2, while in the PP plot, the peak looks stronger for m_1. As LVK accumulates a larger catalog of BBH coalescences, two-dimensional plots such as this will become more relevant as the population model becomes better-resolved.

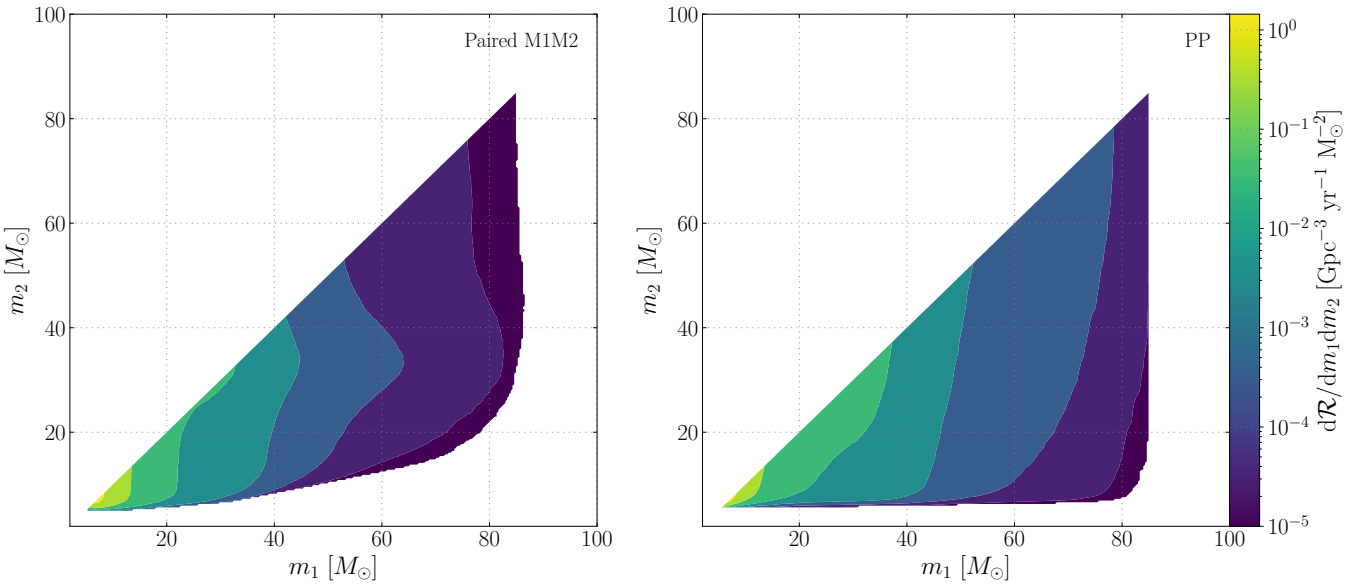

FIG. 8. Median BBH differential merger rate ($\mathrm{d}\mathcal{R}/\mathrm{d}m_1\mathrm{d}m_2$) as a function of both primary mass (m_1) and secondary mass (m_2) for the Paired Primary-Secondary model (left) and the Power Law + Peak model (right). Both plots show the results from GWTC-3.

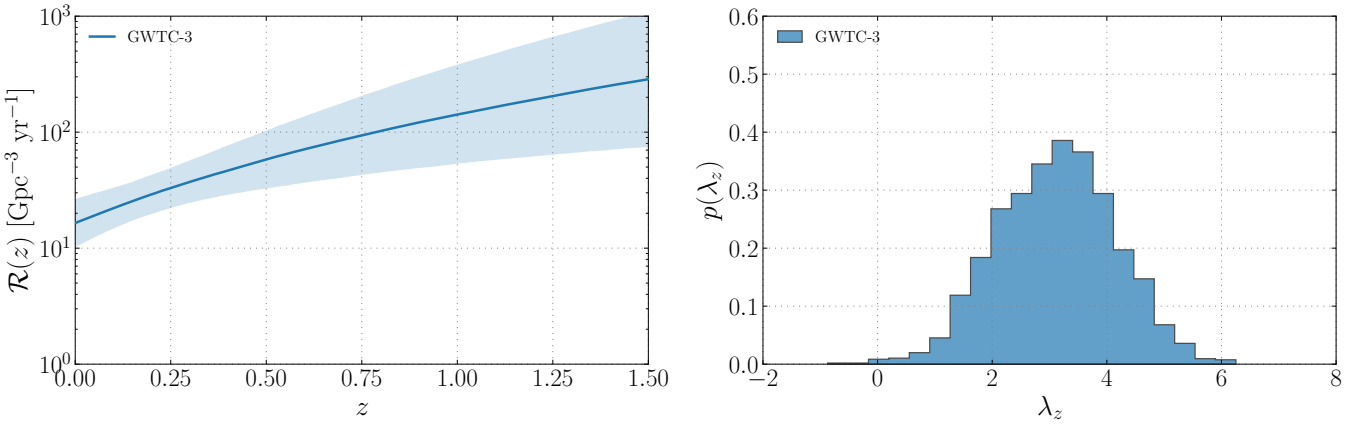

FIG. 9. Redshift evolution of the BBH merger rate, assuming the Power Law Redshift model. The leftmost plot shows the merger rate ($\mathcal{R}$) as a function of redshift (z). The dark blue line shows the median and the light blue shaded region shows the 90% credible interval for GWTC-3. The rightmost plot shows the marginalized posterior distribution for the power-law parameter λ_z, for a model $\mathcal{R}(z) \propto (1+z)^{\lambda_z}$, constrained by the GWTC-3 catalog.

B. Redshift Distribution

1. PLR Model

With GWTC-3, it was shown in Ref. [2] that the low-redshift power-law index for the merger rate was larger than zero with a very high probability. We see the same results here. In Fig. 9, we plot the PLR model results for GWTC-3, as well as the posterior for the power-law spectral index, λ_z. Despite constraining the merger rate reasonably well at low-redshifts, there is approximately a factor of ten between the lower and upper bound of the confidence intervals at $z = 1.5$.

2. MDR Model

A more-accurate redshift model to use should theoretically be the MDR model. However, the GWTC-3 catalog hardly constrains z_p and κ, the hyper-parameters describing the high-redshift merger rate. In Fig. 10, we first (on the left) plot the merger rate according to the MDR model, accounting for the uncertainty in all three hyper-parameters (γ, z_p, and κ). At low-redshift, the behavior is well constrained and is very similar to the PLR model. However, at $z \gtrsim z_p$, the confidence intervals (even at 50% confidence) blow up and give essentially zero information about the merger rate at high redshifts

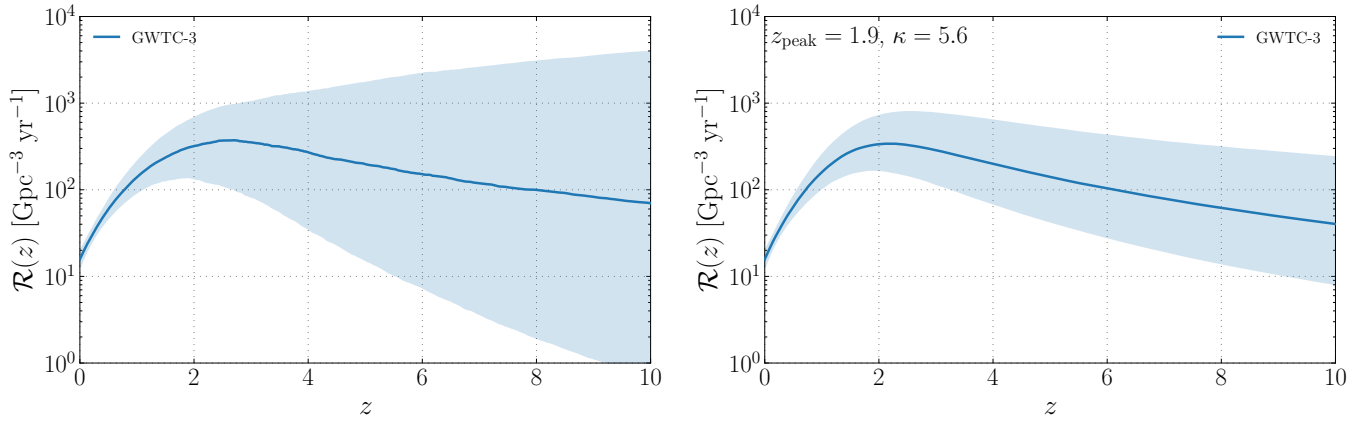

FIG. 10. Redshift evolution of the BBH merger rate, assuming the Madau-Dickinson Redshift model. Both plots show the merger rate ($\mathcal{R}$) as a function of redshift (z), with the light blue shaded regions showing the 50% credible intervals for GWTC-3. Solid black and blue lines show the median. The leftmost plot shows the merger rate, with uncertainty in all three hyper-parameters (γ, z_p, and κ) from the MDR model. The rightmost plot shows the merger rate, with only uncertainty on the low-redshift power-law index, γ; here we fix z_p and κ to specific values indicated in the plot. Note that the rightmost plot is essentially what we use to compute the GWB for the PLR model, marginalizing only over λ_z (which is analogous to γ in the MDR model).

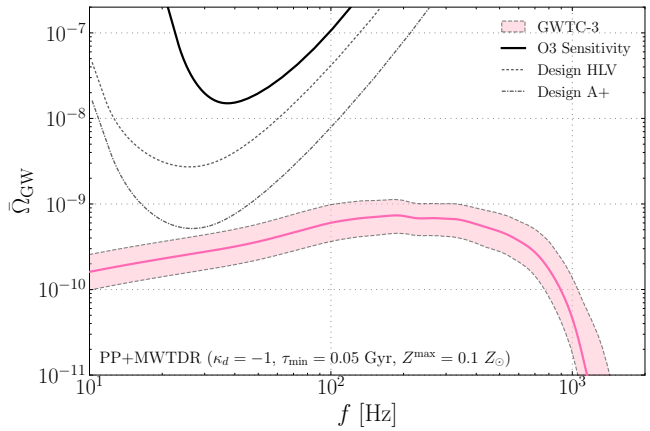

FIG. 11. Dimensionless energy density spectra for the GWB ($\bar{\Omega}_{\rm GW}$) as a function of frequency (f, Hz). This shows a prediction of the GWB from BBHs from the GWTC-3 catalog, assuming the Power Law + Peak mass model and a fixed Metallicity-Weighted Time Delay Redshift model (with a power-law spectral index of -1 for the time delay distribution, a minimum time delay of 0.05 Gyr, and a maximum metallicity of 0.1 $Z_\odot$). The pink shaded region indicates the 90% credible interval for GWTC-3. The pink solid line indicates the GWB from the median-likelihood population model. We denote the current and future LVK sensitivities to an isotropic GWB search with black lines from Ref. [2]. This plot is meant to reproduce Fig. 1.

because LVK has no data at $z \gtrsim 1.5$. This is further compounded upon when we recognize that z_p and κ are *extremely dependent on the prior*; a wider prior would make the uncertainty in merger rate even larger. For this reason, we also plot (on the right) the MDR model with *fixed* z_p and κ. In this merger rate model, the confidence intervals are much less uncertain, and are only different by tenfold at $z = 10$. Currently, the only way to constrain z_p and κ further is to incorporate upper limits on the measured stochastic GWB [6]. However, [30] finds only limited improvements from performing a joint analysis that includes observational limits on the stochastic GWB.

VI. EMPIRICAL GRAVITATIONAL-WAVE BACKGROUND RESULTS

To compute the expected GWB from a population of BBHs, we require a mass model and a redshift model. For this, we should use the MDR (or MWTDR) model rather than the PLR model, since the MDR model does not blow up at high redshift. However, if we plug the PLR spectral index (λ_z) into the low-redshift spectral index for the MDR model (γ), and fix the other MDR parameters to specific values, then we can still visualize the GWB from the PLR model.

In Fig. 11, we show our prediction of the GWB in a similar manner to Fig. 1. In this plot, we marginalize over all parameters of the PP model, as well as the local merger rate. We use a fixed metallicity-weighted time delay distribution of the form $p(\tau) = \tau^{-1}$ with 0.05 Gyr $\leq \tau \leq H_0^{-1}$ and $Z^{\rm max} = 0.1\ Z_\odot$. The main limitation of this plot is that it does not marginalize over uncertainty in the time-delay model nor the metallicity weighting. These distributions are almost completely unknown. However, for future sections, we marginalize over more uncertainties.

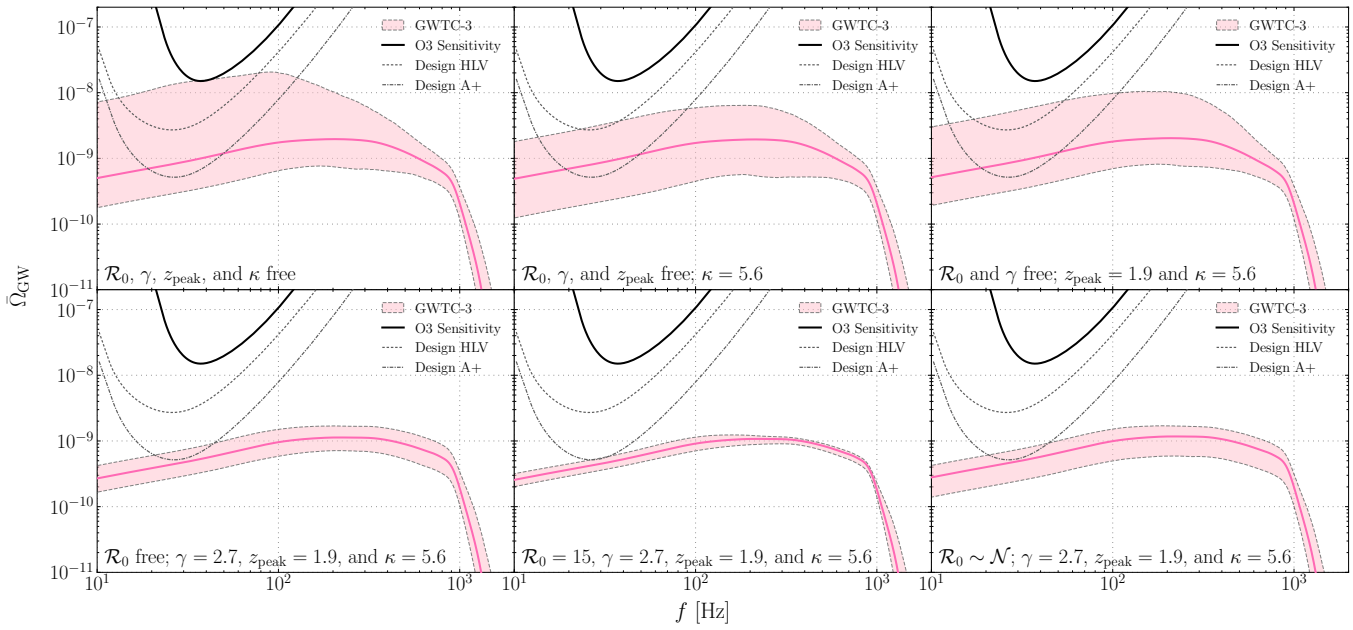

FIG. 12. Dimensionless energy density spectra ($\bar{\Omega}_{\rm GW}$) as a function of frequency (f, Hz). Prediction of the GWB from BBHs from the GWTC-3 catalog, assuming the Power Law + Peak mass model and the Madau-Dickinson Redshift model. The shaded pink regions indicate the 90% credible intervals for GWTC-3. On all plots, we denote the current and future LVK sensitivities to an isotropic GWB search with black lines from Ref. [2]. These lines are the same on all plots, and allow for easy comparison with the eye. Each plot shows the GWB prediction with uncertainty incorporated in a different way: the top left plot marginalizes over all uncertainties; the top middle plot marginalizes over all uncertainty except for κ; the top right plot fixes both $z_{\rm peak}$ and κ; the bottom left plot fixes γ, $z_{\rm peak}$, and κ; the bottom middle plot fixes γ, $z_{\rm peak}$, κ, and the local merger rate $\mathcal{R}_0$; and the bottom right plot fixes γ, $z_{\rm peak}$, and κ, but samples $\mathcal{R}_0$ from a Gaussian distribution with the same mean and standard distribution as the marginalized rate posterior.

A. The Impact of Uncertainty

In Fig. 12, we show the predictions for the GWB using the PP and MDR models, but with varying levels of uncertainty incorporated. Of course, since the BBH catalog cannot constrain the high-redshift behavior of the merger rate, marginalizing over the high-redshift parameters is not particularly relevant, but it is the most conservative estimate of the GWB. It should also be noted that the upper ends of the confidence intervals, when all uncertainty is accounted for (top left plot), are quite high, because the value of κ essentially determines if the merger rate decreases or flattens out (or even increases) after reaching its peak. In particular, only the low-frequency end of the $\bar{\Omega}_{\rm GW}$ curve is increased by marginalizing over κ because high-redshift BBH coalescences are redshifted to lower frequencies. When we move onto the top middle plot in Fig. 12, we can see that the upper end of the confidence intervals decrease accordingly. The median-likelihood population signal remains at approximately the same place because our prior is centered on $\kappa = 5.6$.

When we remove more uncertainty in Fig. 12, the confidence intervals become smaller, as expected. By not marginalizing over $z_{\rm peak}$ (top right plot), the lower end of the confidence intervals rises since the peak redshift can no longer be set to small redshifts. When we remove even γ (bottom left), the confidence intervals shrink even further, and the overall prediction of $\bar{\Omega}_{\rm GW}$ is lower because $\gamma = 2.7$ is smaller than the median γ obtained through population inference. Fixing the local merger rate (bottom middle) makes the uncertainty incredibly small. This shows that the uncertainty in the mass model hyper-parameters does not really affect the GWB prediction; yet, the mass model itself can change the prediction. In the bottom right plot, we include uncertainty in the local merger rate, but instead of deriving that uncertainty from the posterior draws, we instead take the mean and standard deviation from the marginalized rate posterior and plug those into a Gaussian to randomly sample merger rates for each posterior draw. The goal of this is to avoid sampling bias, which can occur when fixing only specific hyper-parameters in the posteriors.

The purpose of analyzing the uncertainties inherent to GWB predictions is to demonstrate the importance of the high-redshift behavior of the merger rate. If the merger rate adheres to our predictions, then the GWB could be detectable by LVK A+. Yet, if the merger rate exhibits unexpected behavior beyond the current redshift sensitivities of LVK, then the GWB could be much larger than currently expected. However, given that the O4a noise sensitivity curve is somewhere between the O3 sensitivity and the Design HLV sensitivity, we will soon be

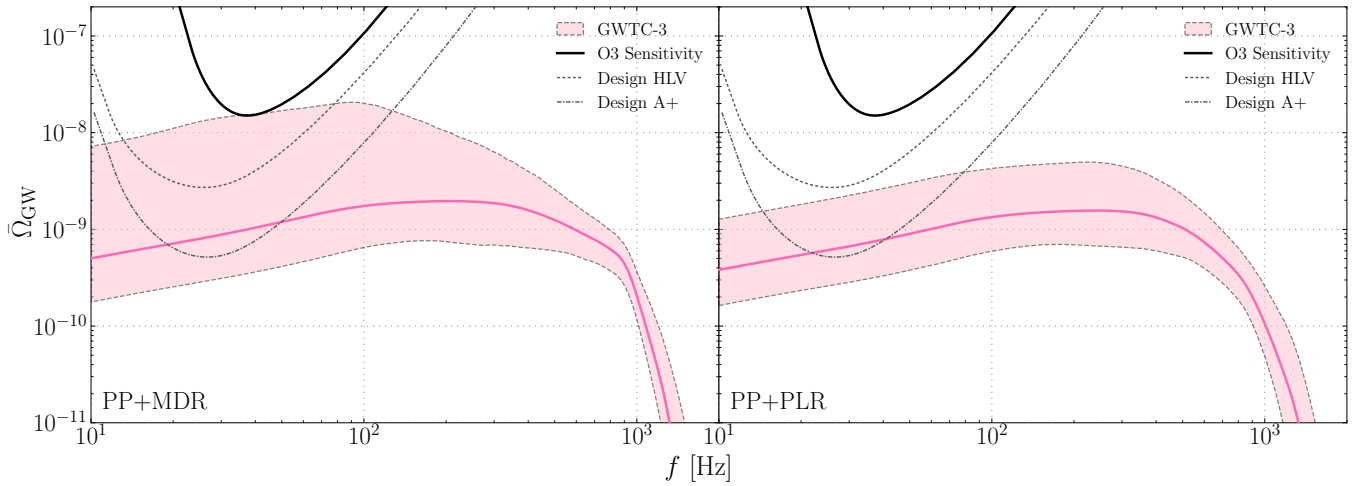

FIG. 13. Dimensionless energy density spectra for the GWB ($\bar{\Omega}_{\rm GW}$) as a function of frequency (f, in Hz). This shows a prediction of the GWB from BBHs from the GWTC-3 catalog, assuming the Power Law + Peak mass model and either the Madau-Dickinson Redshift model (left) or the Power Law Redshift model (right). The pink shaded regions indicate the 90% credible intervals for GWTC-3. The pink solid lines indicate the GWB from the median-likelihood population model. We denote the current and future LVK sensitivities to an isotropic GWB search with black lines from Ref. [2]. For this estimation of the GWB, we marginalize over all relevant uncertainties; the PLR GWB is calculated by plugging in the posterior draws of λ_z into the MDR model and fixing the undetermined values of $z_{\rm peak}$ and κ.

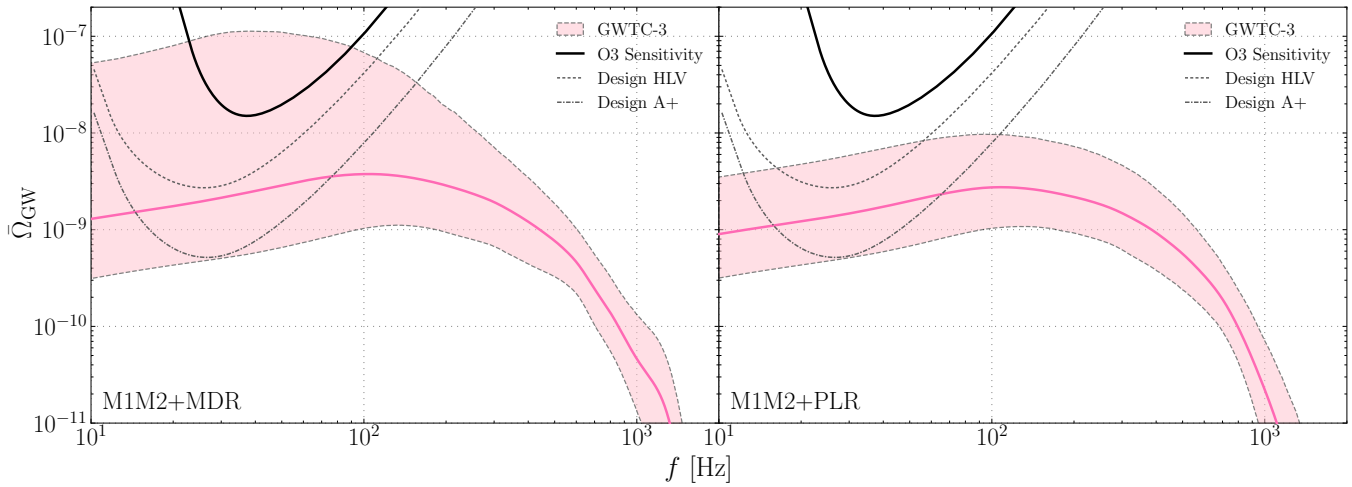

FIG. 14. Same as Fig. 13, but for the Primary-Secondary Identical mass model. This model predicts a higher GWB energy density, especially for lower frequencies, because the M1M2 model lacks a pairing function.

able to rule out some of these exemplary merger rate distributions through a nondetection of the GWB (see Ref. [6]).

B. PP Model

In Fig. 13, we display the our current best prediction of the GWB using the PP model and both the MDR model (left plot) and the PLR model (right). The MDR model tends to overestimate the GWB compared to the PLR model because the high-redshift behavior of the merger rate distribution is not well determined by the GWTC-3 catalog. The only astrophysical assumptions that go into these plots are the models themselves, but all model parameters are determined by the population of BBHs we see (except for fixing $z_{\rm peak}$ and κ to particular values for the case of the PLR model). Given this, it is reassuring to see agreement between these plots and those in Fig. 1, which uses an astrophysically-motivated redshift model.

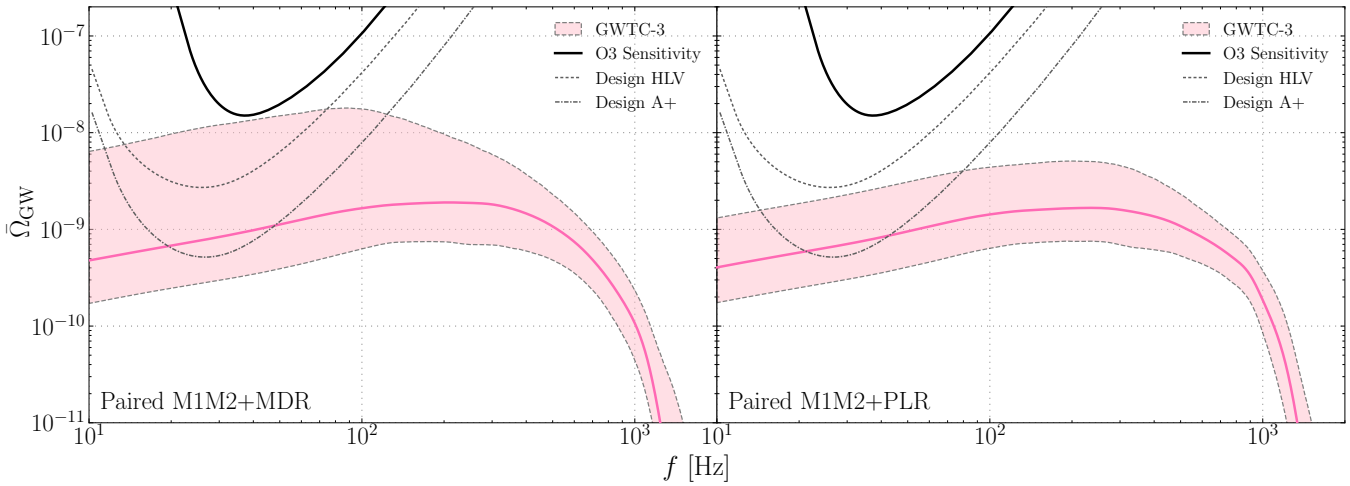

FIG. 15. Same as Fig. 13, but for the Paired Primary-Secondary mass model. These predictions show good agreement with Fig. 13, indicating the necessity of the pairing function.

C. M1M2 Model

In Fig. 14, we display our current best prediction of the GWB assuming the M1M2 model and either the MDR model (left plot) or the PLR model (right). On a first look, it is clear that these results are quite different from Fig. 13. The GWB predicted using the MDR model has extremely large upper uncertainties, and both curves are shifted upwards compared to the PP model by a factor of two or three. Yet, because this model does not include a pairing function, this makes sense.

In the M1M2 model, we assume that both BHs in a BBH system are drawn from the same exact distribution. This allows us to have BHs that are *both* quite large. Because large BHs coalesce at lower frequencies compared to smaller BHs (since their Schwarzschild radii are larger, the orbital frequency must be smaller at the time of merger), the contribution to the GWB from these systems will inherently be biased towards lower frequencies. Another component to consider is that the M1M2 model results in a lower local merger rate, but higher power-law spectral index for the redshift model. This would cause more of the GWB to come from higher redshifts, as there are fewer local mergers implied from such a model. This would then cause the GW signals to be redshifted to lower frequencies. It is also of note that these processes cause the *peak* of the GWB spectrum to occur at lower frequencies. As a whole, this means that the M1M2 model should, and does, predict a larger GWB at low frequencies than the PP model.

The M1M2+PLR models also shift the GWB towards higher energy densities, just like the MDR model, but in this case the uncertainties are not excessive. Given the discrepancy between the GWB predicted by the M1M2 model and the PP model, we can see that the M1M2 model is probably not right.

D. Paired M1M2 Model

In Fig. 15, we show the same analysis as in the previous section, but for the Paired M1M2 model. This computation of the GWB does not run into the limitations from the previous section, due to the usage of a pairing function. In fact, this estimation of the GWB is extremely consistent with Fig. 13. This indicates that the PP model and the Paired M1M2 models are sufficiently similar.

VII. PHENOMENOLOGICAL GRAVITATIONAL-WAVE BACKGROUND RESULTS

Up until now our results have been entirely catalog-driven. We have taken the GWTC-3 catalog, performed population inference to obtain posteriors on our population model hyper-parameters, and fed those posteriors into the GWB calculations to obtain credible intervals on the GWB, given our observed data. Here, we will pivot to a more phenomenological approach, where we will observe the behavior of the GWB under different, physically-motivated, models. In particular, the GWTC can only tell us about the $z \lesssim 1$ behavior of BBH coalescences, so if we want to analyze the possibility of more interesting phenomenon (*e.g.,* a redshift-evolving mass model, one with Pop. III stars, etc.), we must resort to a catalog-free approach. This analysis will be entirely theoretical and not driven by catalogs; hence, we are free to vary parameters which are not constrained by data.

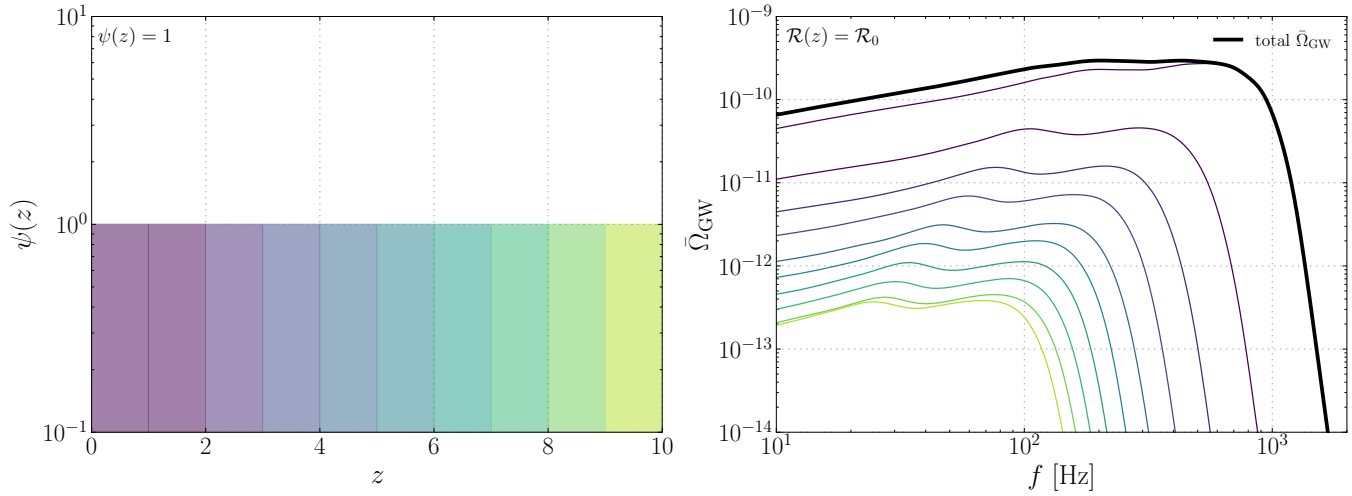

FIG. 16. (left) Normalized merger rate distribution (ψ) as a function of redshift (z), color-coded and split into redshift bins of width one. (right) Dimensionless energy density spectra for the GWB ($\bar{\Omega}_{\rm GW}$) as a function of frequency (f, in Hz). Each colored line in the rightmost plot corresponds to a bin in the leftmost plot, meaning that each colored $\bar{\Omega}_{\rm GW}$ is the GWB caused by a specific range of redshifts. The solid black line on the rightmost plot shows the sum of all the individual binned GWB energy densities. Here we choose the merger rate to be a uniform distribution, just to showcase how the GWB scales with redshift. For these results, we assume the Power Law + Peak mass model with the median-likelihood hyper-parameters, and a local merger rate of $\mathcal{R}_0 = 15$ Gpc^{-3} yr^{-1}.

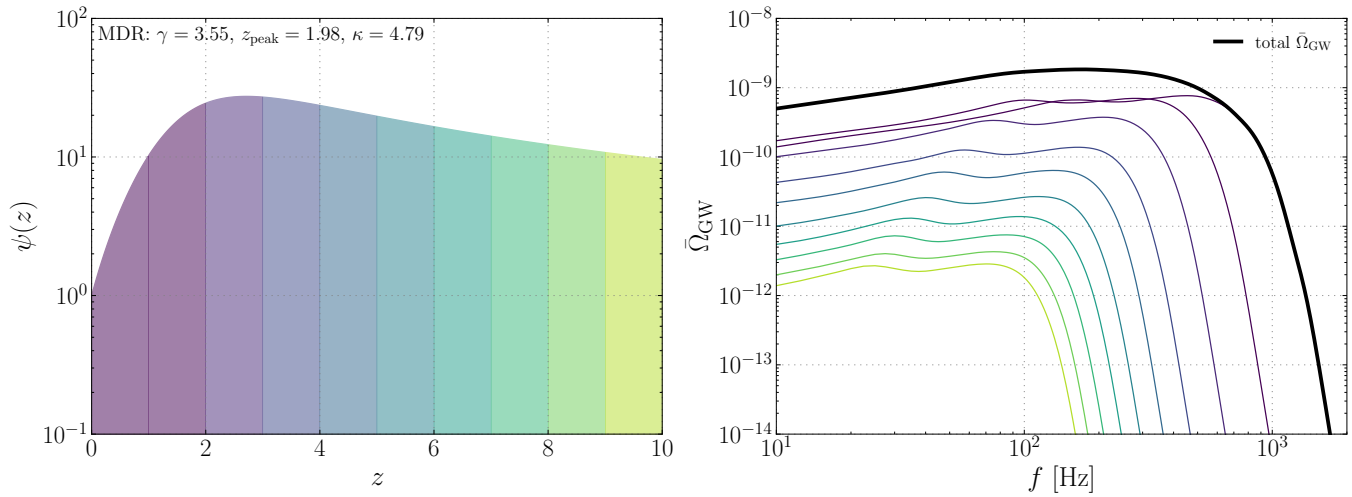

FIG. 17. Same as Fig. 16, but using the Madau-Dickinson Redshift redshift model with its median-likelihood hyper-parameters, and the median-likelihood local merger rate. This is a more realistic decomposition of the GWB energy density.

A. Redshift-Dependence of the Gravitational-Wave Background

We first analyze the way in which the GWB decomposes into different redshifts. To initially characterize this behavior, without the influence of a redshift model, we start with a uniform merger rate distribution. This is plotted in Fig. 16. Here we can see that $0 \leq z \leq 1$ is dominant compared to higher redshifts. It is also clear that the turnoff point of the GWB moves to lower frequencies as redshift increases, as expected.

In Fig. 17, we show the same thing, but with the MDR model. Because of this model, this redshift decomposition is probably more realistic (keeping in mind that we do not know much about what the redshift distribution looks like at high-redshift). We can also see that $0 \leq z \leq 1$ and $1 \leq z \leq 2$ are competitive in terms of the GWB that they each produce. However, from this plot, one can notice that at $z \approx 10$, the predicted GWB is quite low, which could indicate that high-redshift phenomenon such as Pop. III stars could be extremely difficult to observe in the stochastic GWB (unless there happens to be a large time delay).

The competing factors in applying an initially-

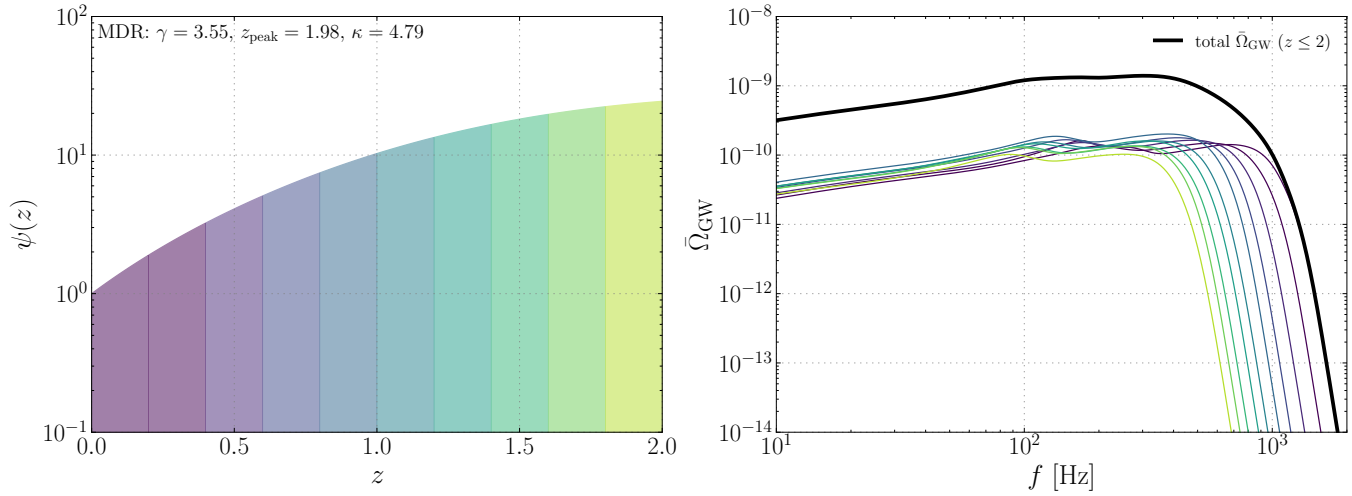

FIG. 18. Same as Fig. 17, but limited to $0 \leq z \leq 2$ with redshift bins of width 0.2. This plot shows how the magnitude of GWB does not monotonically decrease in the presence of an increasing merger rate distribution. Note that the total $\bar{\Omega}_{\rm GW}$ is restricted to $z \leq 2$ so it is lower than in Fig. 17.

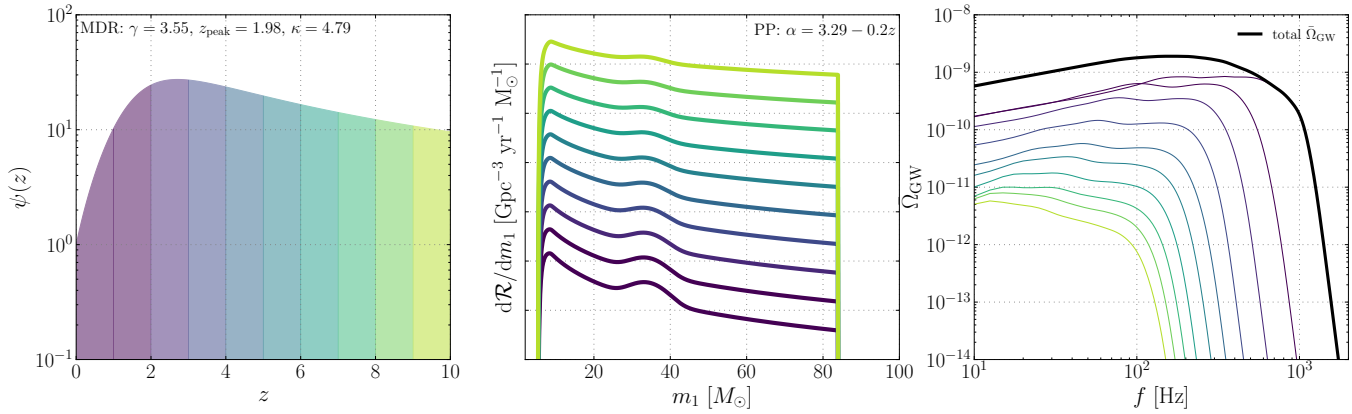

FIG. 19. (left) Normalized merger rate distribution (ψ) as a function of redshift (z), color-coded and split into redshift bins of width one. (middle) differential merger rate ($d\mathcal{R}/dm_1$) as a function of primary mass (m_1) for each of the corresponding redshift bins (arbitrary heights). (right) Dimensionless energy density spectra for the GWB ($\bar{\Omega}_{\rm GW}$) as a function of frequency (f, in Hz). Each colored line in the middle and rightmost plots corresponds to a bin in the leftmost plot, meaning that each colored $\bar{\Omega}_{\rm GW}$ is the GWB caused by a specific range of redshifts. For these results, we assume the Power Law + Peak mass model with the median-likelihood hyper-parameters, the Madau-Dickinson Redshift model with the median-likelihood hyper-parameters, and the median-likelihood local merger rate. For the PP model, we assume that the power-law spectral index evolves as $\alpha(z) = 3.29 - 0.2z$.

increasing merger rate distribution to this redshift decomposition are the decrease in signal magnitude due to larger distance, the increase in comoving volume (so more sources), and the increase in merger rate. This behavior makes it so that our cosmological neighbors are not the dominant contributors to the GWB. In Fig. 18, we show the same plot as in Fig. 17, but for $0 \leq z \leq 2$ with redshift bins of width 0.2. What is immediately clear is that all of these bins contribute roughly equally to the GWB. After zooming in, however, it is apparent that $0 \leq z \leq 0.2$ is not the dominant source of energy density. In fact, it is $0.8 \leq z \leq 1$ that dominates the GWB for $f \lesssim 600$ Hz. This means that in studying the GWB, we cannot always use models derived from our low-redshift neighborhood, as the sources from back when the universe was half its age dominate. It is still a question of how much higher-redshift sources contribute to the GWB.

B. Discrete Redshift Evolution of the Mass Model

A more interesting phenomenon that has been searched for [30, 31] but is not well-constrained is the idea that the mass model evolves with redshift. It is clear that a perfectly-modeled BBH population would include an evolving mass model—hierarchical mergers occur, differ-

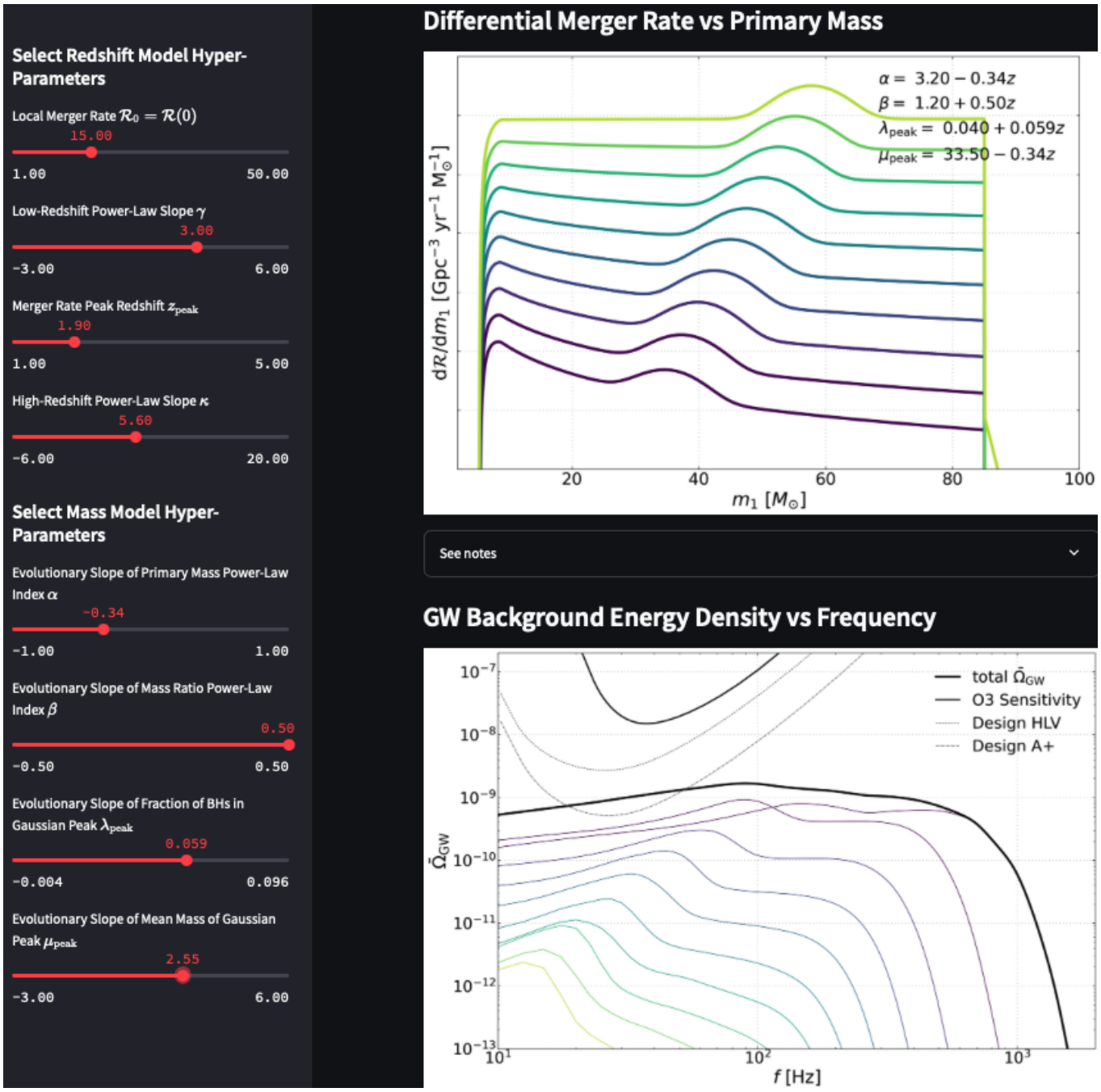

FIG. 20. Screenshot of our "Gravitational-Wave Background Visualizer" app, hosted by Streamlit. The app allows the user to change any of the model hyper-parameters on the left side of the screen, then computes the merger rate ($\mathcal{R}$) as a function of redshift (z, not shown), the differential merger rate ($d\mathcal{R}/dm_1$) as a function of primary mass (m_1, upper plot), and the GWB energy density ($\bar{\Omega}_{GW}$) as a function of frequency (f) (lower plot). The app also includes descriptive notes for the user (under "See notes"). This app is public on Streamlit and the code to reproduce it can be accessed on GitHub.

ent populations of stars (with lower metallicities at higher redshift) behave differently, PPISNe can accumulate over time, etc. Yet, we have not yet observed enough BBH coalescences to constrain this well. But we can still analyze the behavior of the GWB under such a mass-redshift model.

The most basic way to analyze the redshift-dependence of the mass model is by allowing the power-law spectral index of the primary mass model to vary with redshift. So we start with a model that has $\alpha(z) = \alpha_0 + \alpha_z z$. In the context of our histogram redshift intervals, this is easy to implement. So we assign an $\alpha(z)$ to each redshift bin by using the central redshift of the bin. This isn't fully accurate, but it is a good approximation. In

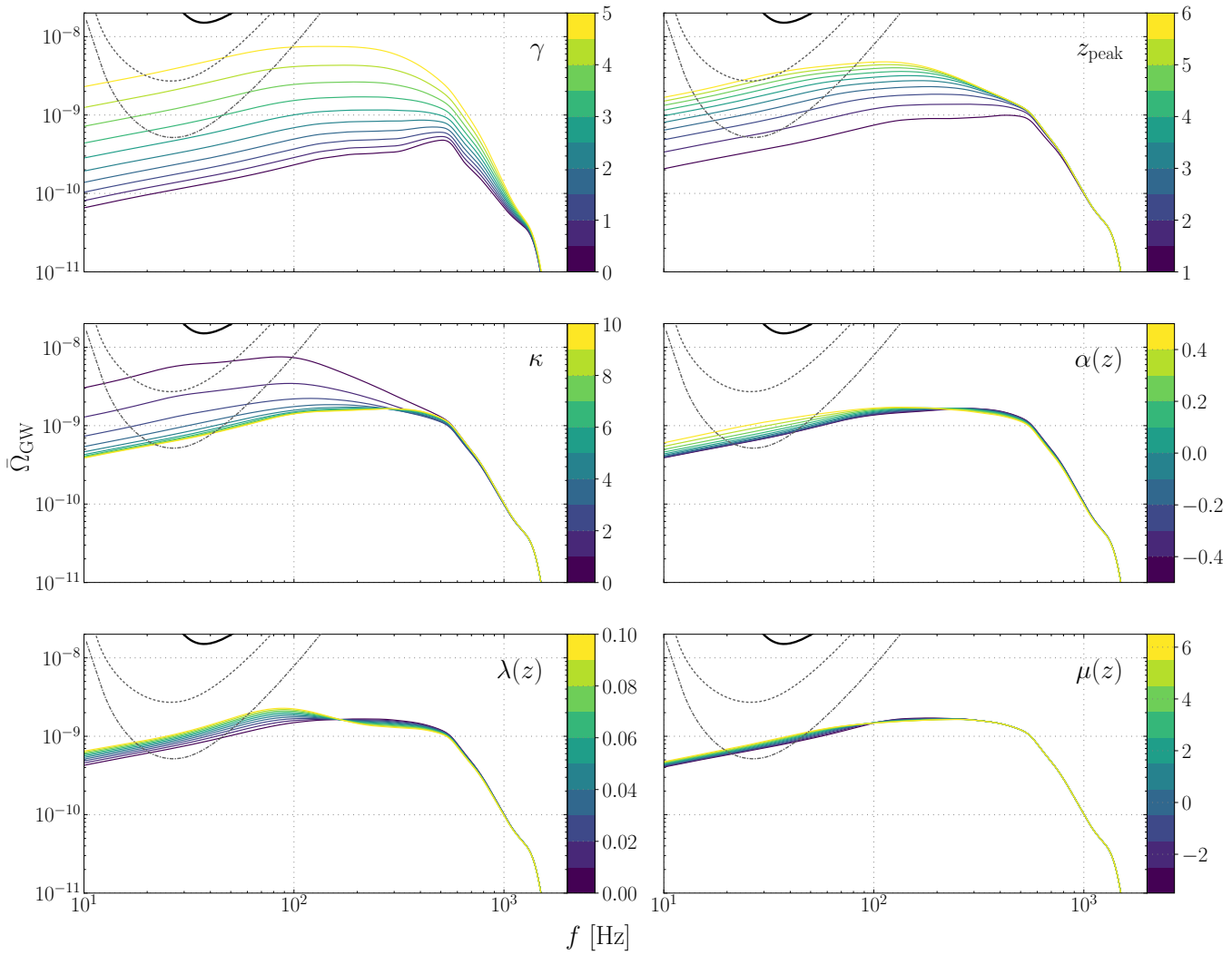

FIG. 21. Dimensionless energy density spectra for the GWB ($\bar{\Omega}_{\rm GW}$) as a function of frequency (f, in Hz), for various values of the hyper-parameters in the Evolving Power Law + Peak model. In each plot, we fix all parameters (to typical values inferred from the population) except for one (denoted in the upper right of each plot), and iterate through that hyper-parameter to show how sensitive the GWB is to that hyper-parameter. As such, the color bars denote the value of each hyper-parameter being iterated over. We use $\lambda(z)$ as a shorthand for $\lambda_{m,z}$, etc. We also show the design sensitivities of O3, O4, and O5 to an isotropic GWB search, as in previous plots. γ is the hyper-parameter describing the low-redshift power-law in the redshift model, $z_{\rm peak}$ is the redshift where the merger rate peaks, and κ describes the high-redshift power-law. $\alpha(z)$, $\lambda(z)$, and $\mu(z)$ describe the linear redshift evolution of the power-law index for primary mass, the area under the Gaussian peak, and the location of the Gaussian peak, respectively.

Fig. 19, we perform this exercise with $\alpha_z = -0.2$. From the plot, it is clear that the mass models flatten out with redshift, and that the GWB energy spectra pivot towards lower frequencies. Such a signature would be difficult to detect: $\alpha_z = -0.2$ increases $\bar{\Omega}_{\rm GW}$ by $10-20\%$ at low frequencies compared to $\alpha_z = +0.2$. This difference could only be detectable with a very accurate measurement of the stochastic GWB.

A more general way to implement redshift-dependence is to allow the primary mass power-law index α, the mass ratio power-law index β, the area of the Gaussian peak λ_m, and the mean of the Gaussian peak μ_m all to vary with redshift. Because of all of the possible ways that these parameters can be varied, we decided that the best way to visualize such phenomenological models is with an app[9]. We show a demonstration of our app in Fig. 20. In this demonstration, we set the mass model to exhibit extreme evolutionary behavior, which causes the redshift slices of the GWB to also exhibit extreme behavior. Nevertheless, the overall GWB is nothing exceptional.

[9] `https://ligo-gwb-viewer.streamlit.app`

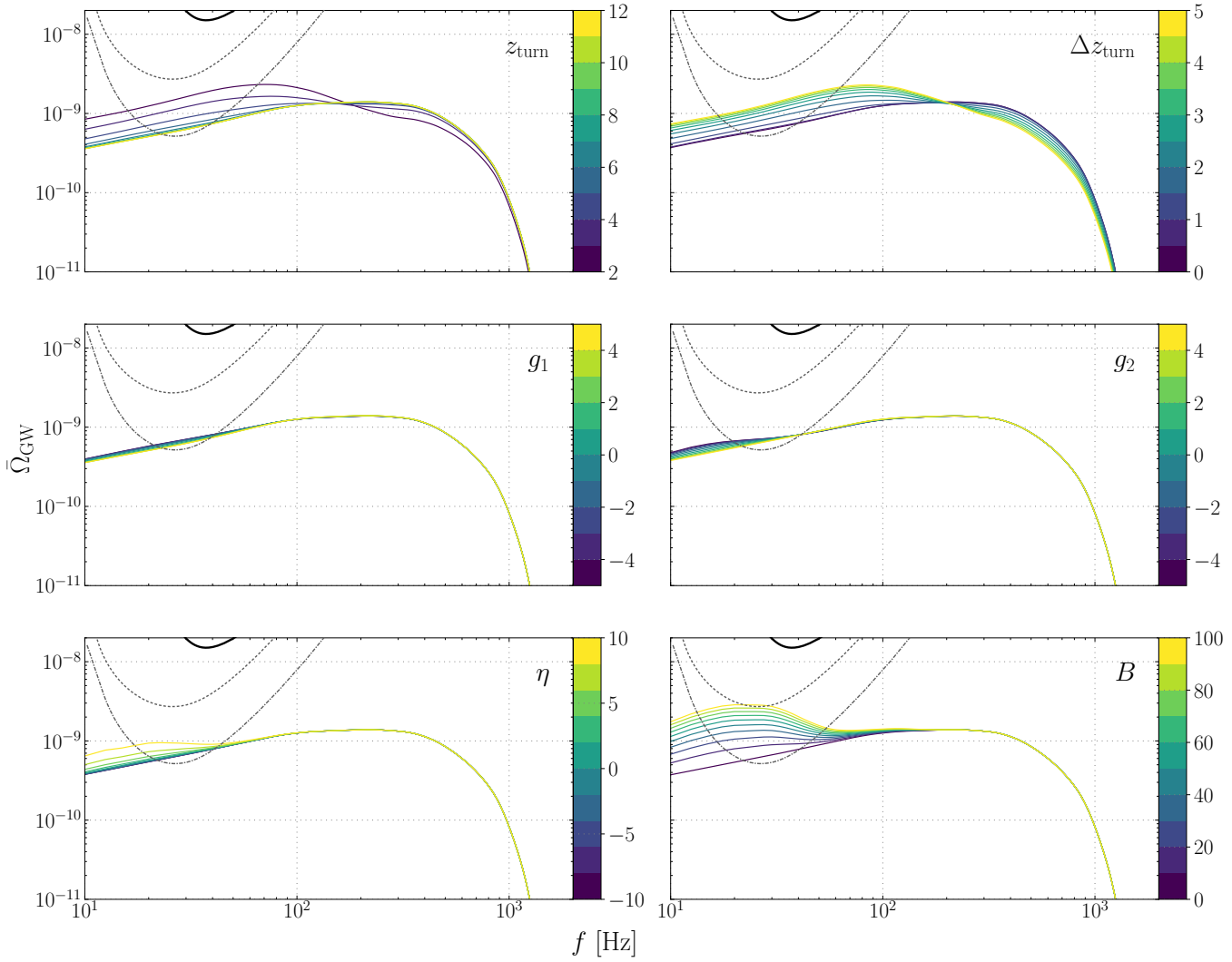

FIG. 22. Same as Fig. 21, but for the Power Peak Globular Cluster model. $z_{\rm turn}$ is the redshift where the mass model is composed of 50% PP and 50% GC. $\Delta z_{\rm turn}$ is the width of that turnoff. g_1 and g_2 are the power-law spectral indices of the mass model at low and high masses, respectively. η is the power-law index on the redshift model for $z > z_{\rm turn}$, and B is the amplification factor applied to the redshift model for $z > z_{\rm turn}$.

C. Continuous Redshift Evolution of the Mass Model

In the previous subsection, we implemented an evolving mass model by discretizing the redshift model into bins, computing the GWB for each bin, then summing up the resulting GWB energy densities. Obviously, a more accurate way to accomplish this is to directly encode redshift evolution into the mass model. We do this for the PP model by creating the EvoPP model.

We first show how the redshift model affects the predicted GWB energy density in Fig. 21 in the top left, top right, and middle left plots. We do this by slicing through a set of values for each hyper-parameter, while holding all other hyper-parameters constant (to typical values). From these, it is clear that the low-redshift power-law index, γ, has the greatest impact. Yet, the peak redshift, $z_{\rm peak}$, has a large influence as well. The hyper-parameter that governs the high-redshift power-law, κ, is less influential, but we can see that if $\kappa \lesssim \gamma$ then the GWB energy density sees a more significant effect. Observing $\kappa \leq \gamma$ in our population would not be completely unreasonable, since Fig. 2 shows that a metallicity-weighted time-delay model can result in a flat—rather than downward-sloping—redshift model at high redshift.

Next, we analyze how the redshift evolution of the mass model affects the predicted GWB. In the middle right, bottom left, and bottom right plots of Fig. 21, we show how varying α_z, $\lambda_{m,z}$, and $\mu_{m,z}$ (linear evolution of the primary mass power-law, area under the Gaussian peak, and mean of the Gaussian peak) affects the GWB. Here we can see that these parameters have a small impact on the GWB magnitude, but they do change the *shape* of

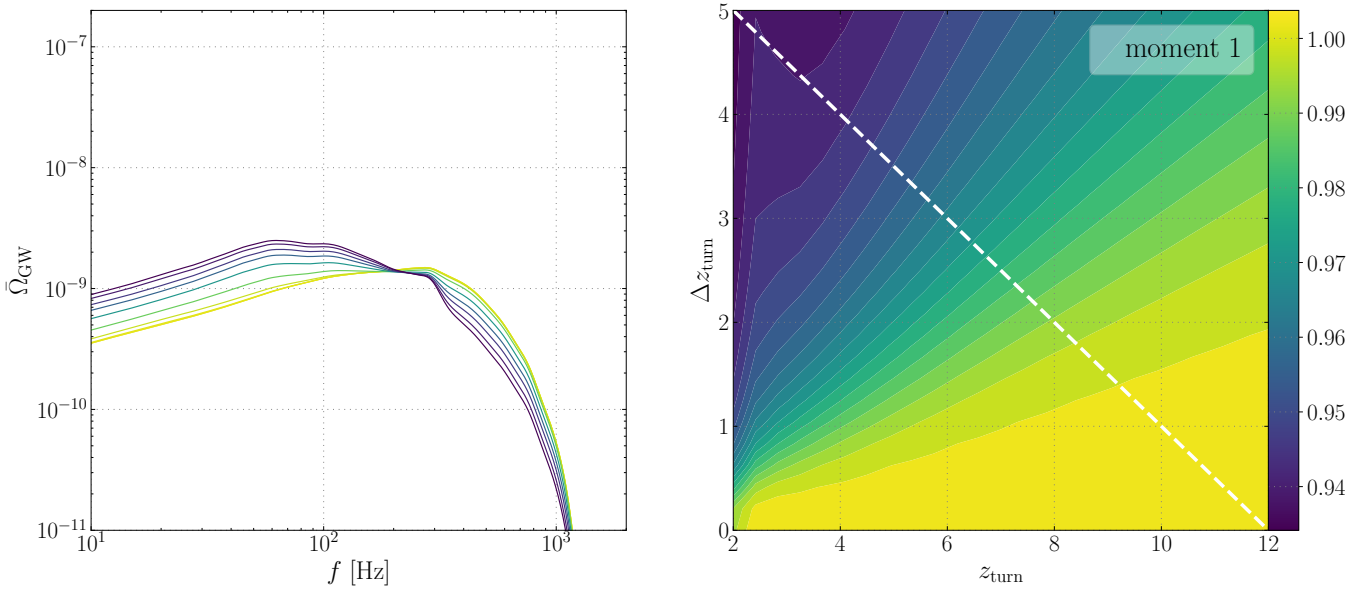

FIG. 23. (left) Dimensionless energy density spectra for the GWB ($\bar{\Omega}_{\rm GW}$) as a function of frequency (f, in Hz), for combinations of ($z_{\rm turn}$, $\Delta z_{\rm turn}$) in the PPGC model. (right) Normalized first moment of the GWB as a function of both turnoff redshift ($z_{\rm turn}$) and turnoff width ($\Delta z_{\rm turn}$). Color indicates the normalized first moment, and the white diagonal like indicates the track of hyper-parameters used in the leftmost plot. Ten sets of hyper-parameters are selected from the white line, and these are then used to plot traces of the GWB energy density on the left.

the GWB. This is most apparent in the $\lambda(z)$ plot, where higher values of $\lambda_{m,z}$ result in a larger magnitude at low-frequencies, but a smaller magnitude at high-frequencies. This is in contrast to the previous three plots, where changing γ, $z_{\rm peak}$, or κ had a large effect on the magnitude of the GWB, but the shape was generally preserved. This hints at the idea that if we can detect the stochastic GWB from the LVK strain data accurately enough, we should be able to make claims about the redshift evolution of the mass model. This is in contrast to the overall magnitude of the GWB, where the effects from γ, $z_{\rm peak}$, or κ are somewhat degenerate with the local merge rate (which sets the absolute normalization).

D. The Power Peak Globular Cluster Model

As seen from the previous subsection, the GWB energy density is not very sensitive to evolution of the mass model (at least for the EvoPP model). This motivates us to perform the same test for a more extreme evolving model. As such, we use the Power Peak Globular Cluster model here.

In Fig. 22, we perform the same analysis as in Fig. 21, but for the hyper-parameters in the PPGC model. From this, we can see that no parameter has a huge effect on the magnitude of the GWB energy density, but some parameters do influence the *shape* quite significantly. In particular, the GWB is most sensitive to $z_{\rm turn}$, $\Delta z_{\rm turn}$, and B. This means that if there is a mid-redshift shift in the population of BBHs, or a high redshift burst of mergers, then LVK could see that in the stochastic GWB. It is also interesting how the $z_{\rm turn}$ and $\Delta z_{\rm turn}$ plots seem to have a stable point at $f \approx 200$ Hz. This could be useful for the detection of an *anisotropic* GWB, which is more sensitive to high frequencies, so could potentially analyze the slope of the GWB energy density at that frequency. Additionally, a burst of high-redshift star formation could be detected through isotropic means, since the hyper-parameter B influences only the lowest frequencies.

Since the shape of the GWB is the most interesting aspect of the GWB that the PPGC model affects, we can analyze the first moment of the GWB under such model changes. Suppose we are given a fiducial GWB energy density spectra $\bar{\Omega}_0$. This can be taken to be the expected GWB under the "typical" population model. We define the normalized first moment of the GWB for a general GWB, $\bar{\Omega}_{\rm GW}$, as

$$M = \frac{\int \mathrm{d}f\ \bar{\Omega}_{\rm GW} f}{\int \mathrm{d}f\ \bar{\Omega}_0 f}. \tag{18}$$

That is, M is the first moment of the GWB, normalized by the first moment of the fiducial GWB. Whereas $\int \mathrm{d}f\ \bar{\Omega}_{\rm GW}$ corresponds to the area under the GWB curve, the first moment can be thought of as the center of mass of the GWB.

In Fig. 23, we show the first moment of the GWB for a range of $z_{\rm turn}$ and $\Delta z_{\rm turn}$ values. We additionally plot a handful of GWB energy densities corresponding to specific $z_{\rm turn}$ and $\Delta z_{\rm turn}$ values. As expected, having a larger $z_{\rm turn}$ increases the first moment because it shifts the GWB towards higher frequencies. Having a smaller

turnoff width additionally increases the first moment because it restricts the influence of the (low-frequency) GC mergers to $z > z_{\rm turn}$; a larger turnoff width allows some GC contributions to occur below the turnoff redshift.

Most combinations of hyper-parameters resulted in uninteresting first moments. However, this form of analysis can be useful for mapping degeneracies and exploring the (hyper) parameter space of our phenomenological models.

VIII. CONCLUSION

We first modelled the population of BBHs that LVK has observed. With this, we modelled the masses using the Power Law + Peak model, the Primary-Secondary Identical model, and the Paired Primary-Secondary Model. We found that the PP and Paired M1M2 models are consistent, but that the M1M2 model is inconsistent with the other models explored in this paper (nature probably does not work in this way). This points at the necessity of implementing a pairing function in the mass model. We found that the Paired M1M2 model is more flexible than the PP model because it directly models both black hole masses, and additionally fits for the *underlying* mass distribution.

We modelled the merger rate distribution of our population of BBHs throughout redshift with the Power Law Redshift model and the Madau-Dickinson Redshift model. These models show consistency at low-redshifts (where LVK makes all of its detections), but—by design—disagree for high-redshifts. In both cases, the credible intervals blow up at high-redshift, meaning that current catalogs do not have the sensitivity to make statements about the high-redshift merger rate.

After modelling the population, we then went on to predict the resulting stochastic GWB energy density. We first used the Metallicity-Weighted Time Delay Redshift model to predict the GWB, marginalizing over uncertainty in the local merger rate and mass model only. After that, we directly marginalized over all uncertainties in the PP model, the M1M2 model, and the Paired M1M2 model. These analyses showed that the true uncertainty in the GWB is extremely large (almost two orders of magnitude). We also found that the PP and Paired M1M2 models make similar GWB predictions, but that the M1M2 model predicts a larger GWB energy density.

After working with current catalogs, we moved on to a catalog-free approach towards analyzing the GWB. In doing so, we analyzed the structure of the GWB energy density, finding that low-redshift BBH events dominate the signal. Yet, with a more realistic merger rate model, all redshifts from $0 \leq z \leq 2$ are comparable. We then analyzed how the GWB changes under a mass model that evolves with redshift, finding that with the typical evolving PP model, the GWB doesn't change significantly. More importantly, we showed that merger rate hyper-parameters such as γ, $z_{\rm peak}$, and κ affect magnitude of the GWB, while evolutionary hyper-parameters (linear evolution of the mass model with redshift) affect the shape of the GWB. We also provided an app to better-visualize the GWB under an evolving mass model.

Finally, we constructed an extreme model that tests the structure of the GWB under unexpected—but physically-motivated—high-redshift behavior. We used this Power Peak Globular Cluster model to show again that evolutionary parameters can significantly influence the shape and the normalized first moment of the GWB energy density spectra. Thus, such signatures may be observed in a future detection of the stochastic GWB signal from stellar-mass BBH coalescences.

In the future, an interesting next step would be to perform population inference using evolving models such as the EvoPP and PPGC models, to then compute the *empirical* GWB predicted from such models. A similar analysis could be performed using the MWTDR model as the redshift model in population inference. If the stochastic GWB is to be accurately interpreted in the future, we must consider a variety of potential population models that could produce similar signals to each other. Eventually, it will be most valuable to use non-parametric models to predict the stochastic GWB.

IX. ACKNOWLEDGEMENTS

JS thanks JG and AW for introducing him to the world of gravitational-wave astronomy and for being incredible mentors. This work was supported by the National Science Foundation Research Experience for Undergraduates (NSF REU) program, the LIGO Laboratory Summer Undergraduate Research Fellowship program (NSF LIGO), and the California Institute of Technology Student-Faculty Programs. This work uses data from LVK observing runs 1–3, but all work is by the authors and does not represent the LVK collaboration as a whole.

[1] B.P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), *Observation of Gravitational Waves from a Binary Black Hole Merger*, Phys. Rev. Lett. **116**, 061102 (2016).

[2] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), *Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3*, Phys. Rev. X **13**, 011048 (2023).

[3] T. Regimbau, *The astrophysical gravitational wave stochastic background*, Res. Astron. Astrophys. **11**, 369 (2011).

[4] G. Agazie *et al.* (The NANOGrav Collaboration), *The NANOGrav 15 yr Data Set: Evidence for a Gravitational-wave Background*, ApJL **951**, L8 (2023).

[5] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), *Upper limits on the isotropic gravitational-wave background from Advanced LIGO and Advanced Virgo's third observing run*, Phys. Rev. D **104**, 022004 (2021).

[6] T. Callister *et al.*, *Shouts and Murmurs: Combining Individual Gravitational-wave Sources with the Stochastic Background to Measure the History of Binary Black Hole Mergers*, ApJL **896**, L32 (2020).

[7] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), *Search for anisotropic gravitational-wave backgrounds using data from Advanced LIGO and Advanced Virgo's first three observing runs*, Phys. Rev. D **104**, 022005 (2021).

[8] M. Pieroni, A. Ricciardone, and E. Barausse, *Detectability and parameter estimation of stellar origin black hole binaries with next generation gravitational wave detectors*, Sci. Rep. **12**, 17940 (2022).

[9] M. Evans *et al.*, *A Horizon Study for Cosmic Explorer: Science, Observatories, and Community*, arXiv e-prints 10.48550/arXiv.2109.09882 (2021).

[10] A. Adamo *et al.*, *The First Billion Years, According to JWST*, arXiv e-prints arXiv:2405.21054 (2024).

[11] P. Madau and M. Dickinson, *Cosmic Star-Formation History*, Annu. Rev. Astron. Astrophys. **52**, 415–86 (2014).

[12] Y. Harikane *et al.*, *GOLDRUSH. IV. Luminosity Functions and Clustering Revealed with ∼4,000,000 Galaxies at z ∼ 2–7: Galaxy–AGN Transition, Star Formation Efficiency, and Implication for Evolution at z>10*, ApJS **259**, 20 (2022).

[13] Y. Harikane *et al.*, *Pure Spectroscopic Constraints on UV Luminosity Functions and Cosmic Star Formation History from 25 Galaxies at z_{spec} = 8.61–13.20 Confirmed with JWST/NIRSpec*, ApJ **960**, 56 (2024).

[14] F. Antonini *et al.*, *Coalescing black hole binaries from globular clusters: mass distributions and comparison to gravitational wave data from GWTC-3*, MNRAS **522**, 466 (2023).

[15] R.S. Klessen and S.C.O. Glover, *The First Stars: Formation, Properties, and Impact*, Annu. Rev. Astron. Astrophys. **61**, 65–130 (2023).

[16] Planck Collaboration, *Planck 2018 results. VI. Cosmological parameters*, A&A **641**, A6 (2020).

[17] E. Thrane and C. Talbot, *An introduction to Bayesian inference in gravitational-wave astronomy: parameter estimation, model selection, and hierarchical models*, PASA **36**, e010 (2019).

[18] C. Talbot and J. Golomb, *Growing pains: understanding the impact of likelihood uncertainty on hierarchical Bayesian inference for gravitational-wave astronomy*, MNRAS **526**, 3495–3503 (2023).

[19] A. Renzini and J. Golomb, *Projections of the uncertainty on the compact binary population background using popstock*, arXiv e-prints (2024).

[20] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run*, Phys. Rev. X **13**, 041039 (2023).

[21] C. Talbot and R. Smith and E. Thrane, *Parallelized inference for gravitational-wave astronomy*, Phys. Rev. D **100**, 043030 (2019).

[22] G. Ashton *et al.*, *BILBY: A User-friendly Bayesian Inference Library for Gravitational-wave Astronomy*, ApJS **241**, 27 (2019).

[23] C. Talbot, *GWPopulation pipe*, Zenodo (2021).

[24] J. Bradbury *et al.*, *JAX: composable transformations of Python+NumPy programs*, GitHub (2018).

[25] C. Talbot and E. Thrane, *Measuring the Binary Black Hole Mass Spectrum with an Astrophysically Motivated Parameterization*, ApJ **856**, 173 (2018).

[26] J. Golomb, M. Isi, and W.M. Farr, *Physical Models for the Astrophysical Population of Black Holes: Application to the Bump in the Mass Distribution of Gravitational Wave Sources*, arXiv e-prints 10.48550/arXiv.2312.03973 (2023).

[27] M. Fishbach and D.E. Holz, *Picky Partners: The Pairing of Component Masses in Binary Black Hole Mergers*, ApJL **891**, L27 (2020).

[28] M. Fishbach and D.E. Holz and W.M. Farr, *Does the Black Hole Merger Rate Evolve with Redshift?*, ApJL **863**, L41 (2018).

[29] K. Turbang *et al.*, *The Metallicity Dependence and Evolutionary Times of Merging Binary Black Holes: Combined Constraints from Individual Gravitational-wave Detections and the Stochastic Background*, ApJ **967**, 142 (2024).

[30] M. Lalleman *et al.*, *Estimating the redshift dependence of the binary black hole population: combining gravitational-wave detections with limits on the stochastic background*, in *The European Physical Society Conference on High Energy Physics* (2024) p. 88.

[31] M. Fishbach *et al.*, *When Are LIGO/Virgo's Big Black Hole Mergers?*, ApJ **912**, 98 (2021).

Appendix A: Mass Models

1. Power Law + Peak

The Power Law + Peak (PP) model consists of a power-law in primary mass from a lower cutoff, $m_{\rm min}$, to an upper cutoff, $m_{\rm max}$, with an additional high-mass Gaussian peak which truncates at a higher mass than the power-law (usually 100 $M_\odot$) [25]. The mass ratio, $q = m_2/m_1$, is modeled as a power-law ranging from $m_{\rm min}/m_1$ to 1, from the definition of mass ratio. Both models are implemented with a smooth turn-on, where a logistic smoothing function tapers the probability density at low masses. See Tab. I for a full description of the model hyper-parameters. Numerically, the PP model for the primary mass is implemented as

$$\begin{aligned} &p(m_1|\alpha, m_{\rm min}, m_{\rm max}, \lambda_m, \mu_m, \sigma_m, \delta_m) \\ &\propto \big[(1-\lambda_m)\mathfrak{B}(m_1|-\alpha, m_{\rm min}, m_{\rm max}) \\ &+ \lambda_m G(m_1|\mu_m, \sigma_m, m_{\rm min}, m_{\rm max,\ gauss})\big] S(m_1|m_{\rm min}, \delta_m) \end{aligned} \tag{A1}$$

Here, p is the probability density for m_1, $\mathfrak{B}$ is a normalized power-law of the form $\mathfrak{B}(x, -\alpha) = Ax^{-\alpha}$ from

TABLE I. Parameters for the Power Law + Peak mass model.

Parameter	Description	Prior
α	Spectral index for the power-law in the primary mass distribution.	$\mathrm{U}(-4, 12)$
β	Spectral index for the power-law in the mass ratio distribution.	$\mathrm{U}(-2, 7)$
$m_{\rm min}$	Minimum BH mass allowed for both the primary mass and mass ratio distributions.	$\mathrm{U}(2.5\ M_\odot, 8\ M_\odot)$
$m_{\rm max}$	Maximum BH mass allowed for the power-law component of primary mass.	$\mathrm{U}(50\ M_\odot, 100\ M_\odot)$
λ_m	Total probability contained within the Gaussian component of primary mass.	$\mathrm{U}(0, 1)$
μ_m	Mean of the Gaussian peak of the primary mass distribution.	$\mathrm{U}(20\ M_\odot, 50\ M_\odot)$
σ_m	Standard deviation of the Gaussian peak of the primary mass distribution.	$\mathrm{U}(1\ M_\odot, 10\ M_\odot)$
δ_m	Width of the smoothing function at low masses.	$\mathrm{U}(0\ M_\odot, 10\ M_\odot)$

$m_{\rm min} \leq x \leq m_{\rm max}$, $G(x, \mu_m, \sigma_m)$ is a truncated Gaussian with mean μ_m, standard deviation σ_m, and edges $m_{\rm min} \leq x \leq m_{\rm max,\ gauss}$, and S is the smoothing function which acts on the low-mass range $m_{\rm min} \leq m_1 \leq \delta_m$. Because of the smoothing function, this probability distribution must be numerically integrated in order to be normalized. Also note that λ_m is the fraction of primary masses that fall within the Gaussian component. The mass ratio distribution is parameterized by

$$p(q|\beta, m_1, m_{\rm min}, \delta_m) \propto \mathfrak{B}(q|\beta, m_{\rm min}/m_1, 1) \times S(q m_1|m_{\rm min}, \delta_m) \qquad \text{(A2)}$$

which also uses the smoothing function (see Ref. [2] for more details on the PP model). Physically, the PP model is motivated by pulsational pair-instability supernovae (PPISNe). In this process, stars with masses $\gtrsim 150\ M_\odot$ leave no black hole remnant because they undergo a runaway thermonuclear explosion [25]. Stars under this mass limit are expected to form black holes of mass $\sim$40 $M_\odot$. Hence, the PPISNe phenomenon is expected to cause a dearth of black holes around $40\ M_\odot \lesssim M \lesssim 250\ M_\odot$ [25]. Physically, this process would manifest itself as a peak in the mass distribution around 40 $M_\odot$, which motivates the Gaussian peak in the PP model. Yet, there is strong evidence that the observed peak in the mass spectrum is not from PPISNe [26]. At low masses, X-ray binaries and population synthesis models predict that the mass distribution peaks above $m_{\rm min}$ (*i.e.,* the distribution requires smoothing, rather than a sharp cutoff) [25]. It should be noted, however, that BBH coalescences inherently produce *larger* black holes, which can throw off this PP model—hierarchical growth is not incorporated, and can cause $m_{\rm max}$ to be larger than the limit imposed by PPISNe.

2. Primary-Secondary Identical

The Primary-Secondary Identical (M1M2) model is similar to the PP model in that it models the BBH mass distribution with a smoothed power-law and a Gaussian peak. Its difference is that it models both the primary mass and secondary mass, rather than the primary mass and mass ratio. The M1M2 model uses an identical distribution for each BH. In essence, this model assumes that all BHs that occur in BBH coalescences are drawn from the same distribution—the fact that the BHs are in a binary has nothing to do with their masses. This is probably wrong because accretion processes add complexity to the physical situation, but it gives some insight into the population of BBHs we are observing. Note that even though we use the same probability distribution for each BH, the *marginalized* distributions for m_1 and m_2 are different because we require that $m_{\rm min} < m_2 < m_1 < m_{\rm max}$. The probability distributions for the M1M2 model are both from Eqn. A1 (replacing m_1 with m_2 for the secondary mass). The priors for this model are listed in Tab. II.

3. Paired Primary-Secondary

A better way to model the primary and secondary masses is to use a pairing function [27]. Thus, the Paired Primary-Secondary (Paired M1M2) model uses a joint probability distribution of

$$p(m_1, m_2|\Lambda) = p_{\rm PP}(m_1|\Lambda) p_{\rm PP}(m_2|\Lambda) f(q, M|\Lambda), \quad \text{(A3)}$$

where $p_{\rm PP}$ denotes the Power Law + Peak probability distribution (Eqn. A1), and $f(q, M|\Lambda)$ is the pairing function, which is a function of mass ratio and total mass. We used a pairing function of

$$f(q, M|\Lambda) = q^{\beta_q} M^{\beta_M} \qquad (q < q_{\rm min}). \qquad \text{(A4)}$$

However, we found that β_M is degenerate with α, so we set a very limited prior on β_M (in alternative to removing it entirely from the pairing function). The priors for the Paired M1M2 model are shown in Tab. III.

Appendix B: Spin Models

The standard spin model we use is the Independent Identical Distribution model (IID). This model assumes that spin magnitudes are drawn from a Beta distribution (the same distribution for each black hole),

TABLE II. Parameters for the Primary-Secondary Identical mass model.

Parameter	Description	Prior
α	Spectral index for the power-law in the mass distribution.	$\mathrm{U}(-4, 12)$
$m_{\min}$	Minimum BH mass allowed for the secondary mass distribution.	$\mathrm{U}(2.5\ M_\odot, 8\ M_\odot)$
$m_{\max}$	Maximum BH mass allowed for the power-law component of primary mass distribution.	$\mathrm{U}(50\ M_\odot, 100\ M_\odot)$
λ_m	Total probability contained within the Gaussian component.	$\mathrm{U}(0, 1)$
μ_m	Mean of the Gaussian peak.	$\mathrm{U}(20\ M_\odot, 50\ M_\odot)$
σ_m	Standard deviation of the Gaussian peak.	$\mathrm{U}(1\ M_\odot, 10\ M_\odot)$
δ_m	Width of the smoothing function at low masses.	$\mathrm{U}(0\ M_\odot, 10\ M_\odot)$

TABLE III. Parameters for the Paired Primary-Secondary mass model.

Parameter	Description	Prior
α	Spectral index for the power-law in the mass distribution.	$\mathrm{U}(-4, 12)$
$m_{\min}$	Minimum BH mass allowed for the secondary mass distribution.	$\mathrm{U}(2.5\ M_\odot, 8\ M_\odot)$
$m_{\max}$	Maximum BH mass allowed for the power-law component of primary mass distribution.	$\mathrm{U}(50\ M_\odot, 100\ M_\odot)$
λ_m	Total probability contained within the Gaussian component.	$\mathrm{U}(0, 1)$
μ_m	Mean of the Gaussian peak.	$\mathrm{U}(20\ M_\odot, 50\ M_\odot)$
σ_m	Standard deviation of the Gaussian peak.	$\mathrm{U}(1\ M_\odot, 10\ M_\odot)$
δ_m	Width of the smoothing function at low masses.	$\mathrm{U}(0\ M_\odot, 10\ M_\odot)$
β_q	Spectral index for the power-law over mass ratio in the pairing function.	$\mathrm{U}(-5, 10)$
β_M	Spectral index for the power-law over total mass in the pairing function.	$\mathrm{U}(-0.1, 0.1)$
$q_{\min}$	Minimum mass ratio allowed in the pairing function.	$\mathrm{U}(0, 0.7)$

and that the spin orientations are drawn from identical distributions that have a Gaussian centered around aligned spins, along with an isotropic (uniform) component. See Tab. IV for a full description of the model hyper-parameters and priors. The spin magnitude (a_1 and a_2) distribution is given by

$$p(a_1, a_2|\alpha_\chi, \beta_\chi, a_{\max}) = \beta(a_1|\alpha_\chi, \beta_\chi, a_{\max}) \times \beta(a_2|\alpha_\chi, \beta_\chi, a_{\max}), \quad \text{(B1)}$$

where $a_{\max}$ is hereafter taken to be 1 because there are no physical limitations that prevent black holes from spinning near the speed of light. The Beta distribution with $a_{\max} = 1$ is given by $\beta(x|\alpha, \beta) \propto x^{\alpha-1}(1-x)^{\beta-1}$. The Beta distribution allows for a flexible set of spin magnitudes; the beta distribution for the BBH population has a spin magnitude distribution with a sharp rise towards a low-spin peak, then a slower decline towards higher-spin black holes. Spin angles are parameterized by $\cos\theta_1$ and $\cos\theta_2$, where θ_1 and θ_2 are the angles between the spin vector and the angular momentum vector of each black hole. The spin orientation probability distribution is given by

$$p(\cos\theta_1, \cos\theta_2|\xi, \sigma) = \frac{1-\xi}{4} + \xi \times G(\cos\theta_1|\mu = 1, \sigma, 1, -1) \times G(\cos\theta_2|\mu = 1, \sigma, 1, -1). \quad \text{(B2)}$$

The first term here is the isotropic component with total probability $(1-\xi)$ (the 4 in the denominator comes in because the integral is over two parameters that each range from -1 to 1). The second term has total probability ξ and is a product of half-Gaussians centered at $\cos\theta = 1$ with standard deviation σ (the Gaussians are also truncated at $\cos\theta = -1$). Physically, the spin orientation model is motivated by a population of black holes in the Gaussian that undergo co-evolution and thus have spins aligned with their orbital angular momentum; the isotropic component is motivated by black holes which undergo dynamical interactions, and come upon each others' orbits randomly, with isotropic spin orientations. See Ref. [2] for more details about the IID model. The IID model is the only spin model considered in this paper, as other spin models use effective and precessing spin as their parameters, which have complicated Jacobians to convert priors. Additionally, the GWB is not very sensitive to the choice of spin model, so the IID model should be sufficient for the purposes of this paper.

Appendix C: Redshift Models

1. Power Law Redshift

The Power Law Redshift model (PLR) parameterizes the BBH redshift distribution as a simple power-law. Obviously, for large redshift, this model is unsustainable. However, at the low redshifts ($z \lesssim 1$) that LVK deals with, the PLR model is perfectly sufficient. For this

TABLE IV. Parameters for the Independent Identical Distribution spin model.

Parameter	Description	Prior
α_χ	First argument of the identical Beta distribution for spin magnitudes.	$\mathrm{U}(0,5)$
β_χ	Second argument of the identical Beta distribution for spin magnitudes.	$\mathrm{U}(0,10)$
ξ	Fraction of probability contained in the Gaussian aligned spin contribution.	$\mathrm{U}(0,1)$
σ	Standard deviation of the Gaussian for aligned spin orientations.	$\mathrm{U}(0.3,5)$

TABLE V. Parameters for the Power Law Redshift model.

Parameter	Description	Prior
λ_z	Spectral index describing redshift model.	$\mathrm{U}(-4,8)$

model, ψ is parameterized as

$$\psi(z|\lambda_z) = (1+z)^{\lambda_z}. \tag{C1}$$

The PLR model is good for modeling the low-redshift behavior of the BBH population, but is not useful for estimating the stochastic GWB from that population, as it blows up for large redshifts. We take a uniform prior on λ_z of $\mathrm{U}(-4,8)$ (Tab. V).

2. Madau-Dickinson Redshift

The Madau-Dickinson Redshift model (MDR) parameterizes the redshift distribution as having a peak at redshift z_p with the following equation [6, 28]:

$$\psi(z|\gamma,\kappa,z_p) = \frac{(1+z)^\gamma}{1+(\frac{1+z}{1+z_p})^\kappa}. \tag{C2}$$

This model is based off the star-formation rate [11] and if the BBH merger rate exactly traced the star-formation rate, we would expect parameters $\gamma = 2.7$, $z_p = 1.9$, and $\kappa = 5.6$. In reality, the redshift distribution should also follow a distribution of time delays, which would change the expected parameters. See Tab. VI for a description of the hyper-parameters and priors of the MDR model. Note that γ and κ are restricted to be ≥ 0 to exclude a redshift distribution that decreases at low-redshift, and also a distribution that explodes at high-redshift.

3. Metallicity-Weighted Time Delay Redshift

While the PLR and MDR models are only slightly astrophysically motivated, the Metallicity-Weighted Time Delay Redshift model (MWTDR) uses some astrophysics to inform its behavior. Note that we can also use consider a time-delay model that does not incorporate metallicity effects, which we occasionally refer to as the Time Delay Redshift model (TDR). The MWTDR model starts with the global star-formation rate (a special case of the MDR model), then assumes that there is some time delay distribution where the lookback time that a BBH coalescence occurs is the time that its stars formed *minus* some time delay between formation and merger. To quantify this, assume there is some function $\mathcal{Z}(t)$ which returns the redshift given the lookback time, and another function $\mathcal{T}(z)$ which returns the lookback time given the redshift of some event. These functions depend on the cosmological model used. Then it is a fact that the redshift of formation, $z_{\rm form}$ can be calculated by summing up the merger time, $t_{\rm merge}$, and the time delay, τ:

$$z_{\rm form} = \mathcal{Z}(\mathcal{T}(z_{\rm merge}) + \tau). \tag{C3}$$

Note that we have to convert the merger redshift into a time in order to add it to the time delay. We also implement the metallicity-weighting function stated in Ref. [29], which describes the fraction of star formation as a function of redshift, given a maximum metallicity:

$$F(Z^{\rm max}, z) = \frac{1}{1.122}\hat{\Gamma}\left[0.84, \left(\frac{Z^{\rm max}}{Z_\odot}\right)^2 10^{0.3z}\right]. \tag{C4}$$

Then the merger rate can be calculated as

$$\mathcal{R}(z) = \mathcal{C}\int \mathrm{d}\tau\ \mathrm{SFR}(z_{\rm form})p(\tau)F(Z^{\rm max}, z_{\rm form}), \tag{C5}$$

where SFR refers to the MDR model with $\gamma = 2.7$, $z_p = 1.9$, and $\kappa = 5.6$, and $p(\tau)$ is the probability distribution of time delays. $\mathcal{C}$ is a constant of proportionality that is influenced by the fraction of formed stars that turn into BHs, the fraction of stars in binary systems, and the fraction of BBHs that coalesce within a Hubble time. Typically, the time delay distribution is taken to be $p(\tau) \propto \tau^{-1}$ [2], but we generalize to $p(\tau) \propto \tau^{\kappa_d}$, where κ_d is some scalar that parameterizes our distribution. We also limit our distribution to have $\tau \geq \tau_{\rm min}$ and $\tau \leq H_0^{-1}$, where $\tau_{\rm min}$ is another parameter and H_0^{-1} is the Hubble time. The time delay distribution is completely unknown, and confounded by dynamical interactions in GCs, so this is only a rough way to infer the redshift distribution. We compute $\psi(z)$ for the MWTDR

TABLE VI. Parameters for the Madau-Dickinson Redshift model.

Parameter	Description	Prior
γ	Spectral index describing the low-redshift power-law.	$\mathrm{U}(0,7)$
z_p	Peak redshift of the distribution.	$\mathrm{U}(0,3)$
κ	Spectral index describing the high-redshift power-law ($\gamma-\kappa$ for $z \gg z_p$).	$\mathrm{U}(0,12)$

TABLE VII. Parameters for the Metallicity-Weighted Time Delay model.

Parameter	Description	Typical Value
κ_d	Spectral index describing probability distribution of time delays.	-1
$\tau_{\min}$	Minimum time delay allowed in the time delay distribution.	0.05 Gyr
$Z^{\max}$	Maximum metallicity used in the fraction of star formation function.	0.1 $Z_\odot$

model using Monte Carlo integration:

$$\psi(z) \propto \frac{1}{N} \sum_{\tau_i \sim p(\tau)} \mathrm{SFR}(z_{\mathrm{form},i}) F(Z^{\max}, z_{\mathrm{form}}). \quad \text{(C6)}$$

We draw the samples τ_i using inverse transform sampling, where we analytically calculate the cumulative distribution function (CDF) for a power-law probability distribution, invert that function, uniformly sample over the interval $[0,1]$, then plug those samples into the inverse CDF to obtain randomly sampled τ_i. To normalize $\psi(z)$ such that $\psi(0)=1$, we can just redo the whole computation once with $z_{\mathrm{merge}}=0$. We summarize the parameters and typical values of the MWTDR model in Tab. VII.

Appendix D: Evolutionary Models

1. The Evolving Power Law + Peak Model

The Evolving Power Law + Peak Model (EvoPP) model is an extension of the PP model, by allowing some of the hyper-parameters in the PP model to linearly change with redshift, *e.g.,* $\xi(z) = \xi_0 + kz$. In this model, we allow α, λ_m, and μ_m to change with redshift:

$$\alpha(z) = \alpha_0 + \alpha_z z \quad \text{(D1)}$$
$$\lambda_m(z) = \lambda_{m,0} + \lambda_{m,z} z \quad \text{(D2)}$$
$$\mu_m(z) = \mu_{m,0} + \mu_{m,z} z. \quad \text{(D3)}$$

The primary computational difference in implementing a mass model that changes with redshift is that one must normalize based on the redshift where the model is being applied. We fully describe the hyper-parameters and list typical values for them in Tab. VIII.

2. The Power Peak Globular Cluster Model

The Power Peak Globular Cluster (PPGC) model is an astrophysically-motivated (see Sec. I) model created to test the behavior of the GWB under a radically-evolving model. This model modifies both the mass model and the redshift model, which we describe below.

a. PPGC Mass Model

The mass model for the PPGC model consists of a Power Law + Peak model and a GC-motivated model, each with their own mixing fraction that changes as a function of redshift. The GC model is described as a double power-law model, *i.e.,*

$$\begin{aligned} &p_{\mathrm{GC}}(m_1|g_1, g_2, m_{\min}, m_{\max}, \mu_m) \\ &\propto \left[\mathcal{C} \times \mathfrak{B}\left(m_1, -g_1, m_{\min}, \frac{\mu_m - m_{\min}}{m_{\max} - m_{\min}}\right)\right. \\ &\left. + \mathfrak{B}\left(m_1, -g_2, \frac{\mu_m - m_{\min}}{m_{\max} - m_{\min}}, m_{\max}\right)\right] \\ &\times S(m_1|m_{\min}, \delta_m), \end{aligned} \quad \text{(D4)}$$

where $\mathcal{C}$ is just a normalization factor to make the overall distribution continuous, $\mathfrak{B}$ is a power-law, and S is a smoothing function. Thus the GC model has a low-mass power-law of g_1 and a high-mass power-law of g_2. We then allow the fraction of mergers that occur from the PP contribution to evolve over redshift as a logistic sigmoid function:

$$\zeta(z|z_{\mathrm{turn}}, \Delta z) = \left[1 + \exp\left(-\frac{z_{\mathrm{turn}} - z}{\Delta z}\right)\right]^{-1}. \quad \text{(D5)}$$

Here, z_{turn} is the time where the PP and GC component have equal contributions to the mass model and Δz is the width of the logistic sigmoid function. Then we can define the overall primary mass distribution for the PPGC

TABLE VIII. Parameters for the Evolutionary Power Law + Peak mass model.

Parameter	Description	Typical Value
α_0	Spectral index for the power-law in the primary mass distribution at $z = 0$.	3.3
α_z	Linear redshift evolution for the power-law in the primary mass distribution.	0
β	Spectral index for the power-law in the mass ratio distribution.	1.2
$m_{\rm min}$	Minimum BH mass allowed for both the primary mass and mass ratio distributions.	5.2 $M_\odot$
$m_{\rm max}$	Maximum BH mass allowed for the power-law component of primary mass.	84 $M_\odot$
$\lambda_{m,0}$	Total probability contained within the Gaussian component of primary mass at $z = 0$.	0.04
$\lambda_{m,z}$	Linear redshift evolution for the fraction of probability contained in the Gaussian peak.	0
$\mu_{m,0}$	Mean of the Gaussian peak of the primary mass distribution at $z = 0$.	33.5
$\mu_{m,z}$	Linear redshift evolution for the mean of the Gaussian peak.	0
σ_m	Standard deviation of the Gaussian peak of the primary mass distribution.	3.8 $M_\odot$
δ_m	Width of the smoothing function at low masses.	4.5 $M_\odot$

TABLE IX. Parameters for the Power Peak Globular Cluster model.

Parameter	Description	Typical Value
	Mass Model Parameters	
α	Spectral index for the power-law in the primary mass PP distribution.	3.3
β	Spectral index for the power-law in the mass ratio distribution.	1.2
$m_{\rm min}$	Minimum BH mass allowed for both the primary mass and mass ratio distributions.	5.2 $M_\odot$
$m_{\rm max}$	Maximum BH mass allowed for the power-law component of primary mass.	84 $M_\odot$
λ_m	Total probability contained within the Gaussian component of primary mass PP distribution.	0.04
μ_m	Mean of the Gaussian peak of the primary mass PP distribution, and location of the break in the primary mass GC distribution.	33.5 $M_\odot$
σ_m	Standard deviation of the Gaussian peak of the primary mass PP distribution.	3.8 $M_\odot$
g_1	Spectral index for the low-mass power-law in the primary mass GC distribution.	-1.5
g_2	Spectral index for the high-mass power-law in the primary mass GC distribution.	3.3
δ_m	Width of the smoothing function at low masses.	4.5 $M_\odot$
	Redshift Model Parameters	
γ	Spectral index describing the low-redshift power-law.	3.3
z_p	Peak redshift of the distribution.	1.9
κ	Spectral index describing the high-redshift power-law ($\gamma - \kappa$ for $z \gg z_p$ and $z < z_{\rm turn}$).	5.6
η	Spectral index describing the very high-redshift power-law ($z > z_{\rm turn}$).	-0.5
B	Amplification factor to the very high-redshift power-law ($z > z_{\rm turn}$).	1
	Evolutionary Parameters	
$z_{\rm turn}$	Redshift where the mass model changes to GC-dominated and the redshift model changes slope.	6
$\Delta z_{\rm turn}$	Width of the logistic sigmoid function centered around $z_{\rm turn}$.	0.8

model to be

$$\begin{aligned} &p(m_1|\alpha,\beta,m_{\rm min},m_{\rm max},\lambda_m,\mu_m,\sigma_m,g_1,g_2,z_{\rm turn},\Delta z,\delta_m) \\ &\propto \Big\{p_{\rm PP}(m_1|\alpha,m_{\rm min},m_{\rm max},\lambda_m,\mu_m,\sigma_m,\delta_m)\times\zeta(z|z_{\rm turn},\Delta z) \\ &+ p_{\rm GC}(m_1|g_1,g_2,m_{\rm min},m_{\rm max},\mu_m)\times[1-\zeta(z|z_{\rm turn},\Delta z)]\Big\} \\ &\times p_{\rm PP}(q|\beta,m_1,m_{\rm min},\delta_m). \end{aligned} \tag{D6}$$

This consists of the PP contribution and the GC contribution—both smoothed at low masses—multiplied by the PP mass ratio model (smoothed as well).

b. PPGC Redshift Model

The redshift model we define for the PPGC model is very similar to the MDR redshift model, but it allows for more flexibility for the high-redshift merger rate. This is roughly defined as

$$\begin{aligned} &\psi(z|\gamma,z_{\rm peak},\kappa,\eta,B) \\ &= \psi_{\rm MDR}(z|\gamma,z_{\rm peak},\kappa)\times H(z_{\rm turn}-z) \\ &+ \mathcal{C}\times B(1+z)^\eta H(z-z_{\rm turn}), \end{aligned} \tag{D7}$$

where $\psi_{\rm MDR}$ is the normalized merger rate from the MDR model, H is the Heaviside step function, $\mathcal{C}$ is a constant that makes the curve continuous when $B = 1$, $z_{\rm turn}$ is the redshift where the merger rate changes to a different power-law, η is the spectral index of that power-law, and B is an amplification factor on the high-redshift component of the model. We do not use this model to perform population inference because its unique features appear beyond the redshift limits of LVK, but we describe the parameters and typical values of them in Tab. IX.