

Detection of LIGO-Virgo binary black holes in the pair-instability mass gapB. O'Brien¹, M. Szczepańczyk¹, V. Gayathri¹, I. Bartos¹, G. Vedovato^{2,3},
G. Prodi^{4,5}, G. Mitselmakher¹, and S. Klimentko¹¹*Department of Physics, University of Florida, PO Box 118440, Gainesville, Florida 32611-8440, USA*²*Università di Padova, Dipartimento di Fisica e Astronomia, I-35131 Padova, Italy*³*INFN, Sezione di Padova, I-35131 Padova, Italy*⁴*Università di Trento, Dipartimento di Fisica, I-38123 Povo, Trento, Italy*⁵*INFN, Trento Institute for Fundamental Physics and Applications, I-38123 Povo, Trento, Italy*

(Received 2 June 2021; accepted 31 August 2021; published 4 October 2021)

By probing the population of binary black hole (BBH) mergers detected by LIGO-Virgo, we can infer properties about the underlying black hole formation channels. A mechanism known as pair-instability (PI) supernova is expected to prevent the formation of black holes from stellar collapse with mass greater than $\sim 40\text{--}65 M_{\odot}$ and less than $\sim 120 M_{\odot}$. Any BBH merger detected by LIGO-Virgo with a component black hole in this so-called PI mass gap likely originated from an alternative formation channel. Here, we firmly establish GW190521 as an outlier to the stellar-mass BBH population if the PI mass gap begins at or below $65 M_{\odot}$. In addition, for a PI lower boundary of $40\text{--}50 M_{\odot}$, we find it unlikely that the remaining distribution of detected BBH events, excluding GW190521, is consistent with the stellar-mass BBH population.

DOI: [10.1103/PhysRevD.104.082003](https://doi.org/10.1103/PhysRevD.104.082003)**I. INTRODUCTION**

During the first two observing runs of Advanced LIGO and Advanced Virgo (O1 and O2), 11 gravitational wave (GW) signals were identified by the LIGO-Virgo Collaboration [1,2] including 10 from binary black hole (BBH) mergers [3]. Substantial improvements in detector sensitivity prior to the first half of the third observing run (O3a) led to the detection of 39 new GW event candidates, at least 36 of which are consistent with BBH mergers [4]. This growing population of BBH mergers allows us to probe underlying black hole formation channels [5,6].

Black holes can form directly through stellar collapse, mergers, accretion, or through collapse of dense gas in the early Universe (primordial black holes). Binary systems consisting of black holes which formed from stellar collapse are referred to as stellar-mass BBHs. In the case of stellar evolution, a mechanism known as pair-instability (PI) supernovae is expected to prevent the formation of heavier black holes [7–11]. In PI, the production of positron-electron pairs causes a reduction in radiation pressure. This in turn causes the core of the star to contract, raising the temperature and increasing the production of positron-electron pairs in a runaway process. Any star with a helium core mass $32 M_{\odot} \lesssim M_{\text{HE}} \lesssim 64 M_{\odot}$ undergoes pulsational pair-instability (PPI). In PPI, the stellar envelope is removed and the mass of the star is reduced until the star reaches a stable state [7,12–14]. Any star with a helium core mass $64 M_{\odot} \lesssim M_{\text{HE}} \lesssim 135 M_{\odot}$ undergoes PI supernova which blasts the star apart, leaving no

remnant compact object [7,15,16]. More massive stars are predicted to collapse directly to black holes [16–18].

The PI process is heavily dependent on stellar metallicity [19,20]. A metal-poor star ($Z \leq 3 \times 10^{-4}$) may collapse into a black hole with mass up to $\sim 65 M_{\odot}$ if the star's hydrogen envelope falls onto the compact object and up to $\sim 45 M_{\odot}$ otherwise [21]. Compared to a low metallicity star, a metal-rich star is expected to collapse into a lower mass compact object [8,16,22–27]. Overall, this mechanism results in an absence of black holes in the mass spectrum known as the PI mass gap. Some theoretical uncertainties regarding the lower boundary remain [28–32], however, $65 M_{\odot}$ is likely a conservative lower limit [8–11,16,33,34]. The upper boundary of the PI mass gap is more certain and is believed to be $\sim 120 M_{\odot}$ [8,9,11].

Black holes can exceed the PI mass limit if they are formed from the previous merger of smaller black holes. This so-called hierarchical merger scenario can occur through dynamical encounters in dense stellar clusters such as galactic nuclei, globular clusters [35–46], or nuclear star clusters [47,48]. Hierarchical mergers can also form through gas-capture in the disks of active galactic nuclei (AGNs) [49–55]. Alternatively, black holes can exceed the PI mass limit by substantial accretion, which is possible for black holes in AGN disks [56,57]. Black holes can also form within the PI mass gap when young stars within a dense young star cluster merge in a runaway collision process [58–60].

GW190521 is the heaviest BBH merger detected to date and marks the first direct detection of an intermediate mass black hole [61]. Reference [34] estimates the probability for at least one of the component black holes to lie within 65–120 M_{\odot} to be 99.0%, assuming a binary system with quasicircular orbit. This probability increases when the analysis was repeated after taking into account the population of detected binary systems [6]. Hence, GW190521 is an appealing candidate for alternative formation scenarios—a prospect which has already been explored in various papers [62–69]. We caution, however, that analyzing the features of a potential outlier BBH event outside the context of the entire population of detected events can be misleading (a notion previously examined by Fishbach *et al.* in Ref. [70]).

Here, we investigate whether GW190521 or any other BBH merger is an outlier to the stellar-mass BBH population. Distinguishing between separate alternative formation channels is outside the scope of this work. We examine the BBH events detected by Coherent WaveBurst (cWB), a model waveform independent pipeline used to search for GW signals [71,72], and compare this sample of detected events against models of the stellar-mass BBH population. In the method presented below, we add simulated BBH waveforms (injections) directly into the GW detector strain data and recover them with the cWB pipeline. This allows us to directly account for detector noise when estimating the significance of potential outliers. Ultimately, we seek to answer two fundamental questions: (i) accounting for statistical fluctuations, are there any outliers to the stellar-mass BBH population? and (ii) is the distribution of detected BBH events consistent with the stellar-mass BBH population?

This paper is organized as follows. In Sec. II, we familiarize the reader with cWB—a pipeline that uses minimal assumptions to identify GW transient signals. Next, we specify the stellar-mass BBH population models and injections used in this study. Then, we define our test statistic and describe our statistical procedure used to classify outliers. In Sec. III, we present our results for various possible models of the stellar-mass BBH population. Finally, in Sec. IV, we state the conclusions of this study and briefly discuss our future work.

II. METHOD

A. Coherent waveburst

CWB is a search algorithm that identifies GW signals in the LIGO-Virgo detector data without directly using a waveform model [71,72]. The signal detection process is performed in the time-frequency domain using the Wilson Daubechies Meyer wavelet transform [73]. Here, the algorithm identifies wavelets inconsistent with detector noise and assembles them into clusters. If a cluster of wavelets is coincident in time and frequency across the

detector network, then an event trigger is generated. For each event trigger, summary statistics are calculated which describe the time-frequency evolution (f_0 , ΔT_s , ΔF_s , \mathcal{M} , F_M , e_M), signal strength and coherence (SNR_{net} , c_c , n_f), and the likelihood fit (E_c/L). Additional statistics are used to estimate the number of cycles in the reconstructed waveform (Q_0 , Q_1). The summary statistics used in this study are defined in Appendix C.

Through O1, O2, and O3a, cWB has contributed to the detection of 22 BBH merger candidates (displayed in Table I). Among these detections is GW190521, which cWB played a crucial role in identifying [61,74]. In this

TABLE I. List of 22 BBH merger event candidates observed by cWB during the O1, O2, and O3a observing runs. Summary statistics, including the central frequency f_0 and network signal-to-noise ratio SNR_{net} , are reconstructed by the cWB pipeline and fed into XGBoost to estimate the source frame total mass M_{cWB} (described in Sec. II D). Events in the table are ordered according to M_{cWB} . The estimated source frame mass M_{cWB} can be compared against the estimated source frame mass M_{GWTC} reported in the LIGO-Virgo Transient Catalogs [3,4].

Event	SNR_{net}	f_0 [Hz]	M_{cWB} [M_{\odot}]	M_{GWTC} [M_{\odot}]
GW190521	14.4	57.8	193.2	$157.9^{+37.5}_{-20.9}$
GW190706_222641	12.7	74.1	139.2	$101.1^{+18.0}_{-13.5}$
GW190602_175927	11.1	76.9	126.8	$114.0^{+18.4}_{-15.5}$
GW190519_153544	14.0	89.2	115.5	$104.2^{+14.5}_{-15.0}$
GW170729	10.2	81.8	110.5	$84.4^{+15.8}_{-11.1}$
GW190701_203306	10.2	93.3	104.3	$94.1^{+11.6}_{-9.2}$
GW190521_074359	24.7	91.4	92.9	$74.4^{+6.9}_{-2.4}$
GW170814	17.2	115.4	90.2	$55.9^{+3.4}_{-2.6}$
GW190421_213856	9.3	87.2	84.6	$71.7^{+12.5}_{-8.6}$
GW190727_060333	11.4	103.5	81.2	$65.8^{+10.8}_{-7.3}$
GW190503_185404	11.5	109.8	78.8	$71.3^{+9.3}_{-8.1}$
GW170823	10.8	114.2	73.5	$68.7^{+10.8}_{-8.1}$
GW150914	25.2	118.7	68.6	$66.1^{+3.8}_{-3.3}$
GW190915_235702	12.3	119.9	67.1	$59.5^{+7.6}_{-6.2}$
GW170104	13.0	147.0	65.3	$51.0^{+5.3}_{-4.1}$
GW190828_063405	16.6	103.5	60.9	$57.5^{+7.4}_{-4.5}$
GW190408_181802	14.8	117.3	52.2	$42.9^{+4.0}_{-2.9}$
GW151226	11.9	120.6	45.8	$21.5^{+6.2}_{-1.5}$
GW190412	19.7	108.5	45.6	$38.4^{+3.8}_{-3.7}$
GW190517_055101	10.7	133.4	41.8	$61.9^{+10.1}_{-9.6}$
GW190512_180714	10.7	125.7	33.1	$35.7^{+3.8}_{-3.4}$
GW170608	14.1	116.9	17.4	$18.6^{+3.2}_{-0.7}$

work, we compare these BBH detections against predictions made by the stellar-mass BBH population. Simulated waveforms representative of the stellar-mass BBH population are injected into the GW detector data and recovered with cWB. We restrict our analysis to events detected by cWB with a false alarm rate (FAR) less than 1 yr^{-1} to maintain a high purity sample set. Although the inclusion of Virgo generally improves source property reconstruction, we simplify our analysis by considering only the Hanford-Livingston detector network.

B. Stellar-mass BBH population models

We examine two phenomenological mass models presented in Ref. [6]: the Broken Power Law model [75] and the Power Law + Peak model [76]. Each population model is described by a set of hyperparameters whose values are estimated by applying hierarchical Bayesian inference on the full population of BBH merger candidates detected by LIGO-Virgo [77–79]. Below is a short description of each model:

- (i) Broken Power Law: The primary mass distribution is characterized by a gentle sloping power law at low mass designed to capture the observed Salpeter initial mass function [80,81] and a second steep sloping power law at high mass which targets a mass tapering of the distribution. The mass ratio is given as a power law that favors equal mass binaries [82].
- (ii) Power Law + Peak: The primary mass distribution is characterized by a power law component at low mass and wide Gaussian component at high masses which models a build-up of black holes due to PPI supernovae [83,84]. Similar to the previous model, the mass ratio is given as a power law favoring equal mass binaries.

The left plot in Fig. 1 displays the astrophysical distribution for the primary mass m_1 predicted by each model, averaged over uncertainty. These mass models describe the complete astrophysical distribution of BBHs and do not distinguish between separate formation channels. To isolate the stellar-mass BBH population, we apply

a theoretically motivated cutoff on the primary mass to model the PI mass gap lower boundary. We assume that no stellar-mass BBHs form inside of the PI gap. To account for the uncertainty surrounding the PI mass limit, we consider three cases: (i) $m_1 < 65 M_\odot$, (ii) $m_1 < 50 M_\odot$, and (iii) $m_1 < 40 M_\odot$. For brevity, we abbreviate each model and add a subscript indicating the primary mass cutoff, e.g., BPL₆₅ indicates the Broken Power Law model with a primary mass cutoff of $m_1 < 65 M_\odot$.

C. Injections

An essential part of this work is that we add injections directly to GW detector strain data. This allows us to compare our detected BBH events against the population models, accounting for selection bias. We use the SEOBNRv4 waveform approximant [85] and inject simulated waveforms uniformly in time throughout the O1-O3a observing runs. In the case of O1 and O2, we use the public GW strain data available through the Gravitational Wave Open Science Center (GWOSC) [86]. At the time of this publication, the GW strain data for the complete O3a observing run was not publicly available. Instead, we used the segments of data available from the GWTC-2 GWOSC data release [86] to emulate the O3a sensitivity, and we repeated injection analysis on these data segments until realizing the full O3a observation time.

The component masses for each simulated BBH event are drawn from the tested stellar-mass model under consideration (see Sec. II B). Component black hole spins are drawn from a uniform aligned spin distribution. This spin distribution is theoretically motivated to favor isolated binaries [87–92]. The binary source distance is drawn assuming uniform density in comoving volume up to redshift $z = 2$ [93]. Other extrinsic binary parameters including sky location and inclination angle are randomly selected.

We recover the simulated BBH events using the cWB pipeline and select events with a FAR less than 1 yr^{-1} . The right plot in Fig. 1 displays the distribution of detected events for each population model. By comparison to the

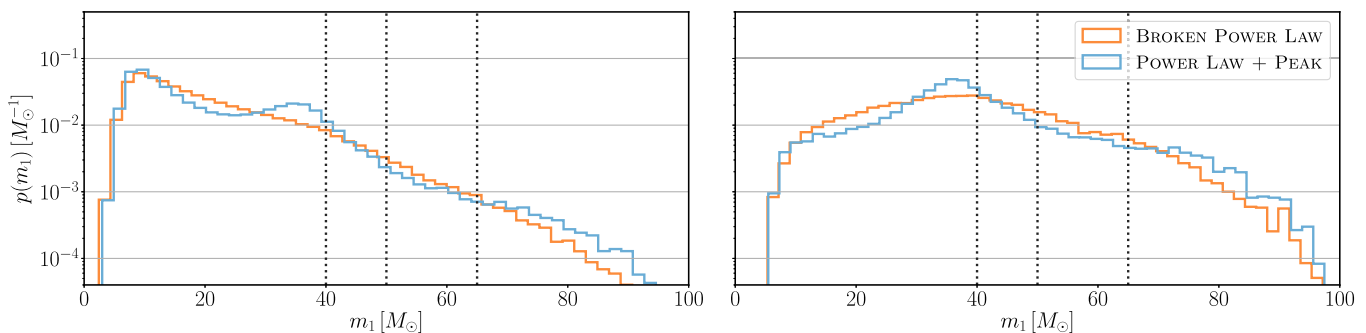


FIG. 1. Expected probability distribution of source frame primary mass m_1 , averaged over uncertainty, for the Broken Power Law (orange) and Power Law + Peak (blue) models: (left) astrophysical distributions, (right) distributions of events detected by cWB with FAR less than 1 yr^{-1} . Dotted vertical lines indicate tested values of the PI mass limit: $40 M_\odot$, $50 M_\odot$, and $65 M_\odot$.

astrophysical distributions, the detected distributions favor more massive BBH mergers due to selection effects caused by the ground-based detectors and the cWB search pipeline.

D. Population outlier test statistic

In this study, we require a test statistic that captures the characteristics of an outlier to the stellar-mass BBH population. One logical statistic to consider is the source frame primary mass m_1 , since stellar-mass BBH events are not expected to have m_1 inside the PI mass gap. Ideally, one could use the source frame primary mass m_1 estimated via Bayesian inference. However, since we inject and recover $\mathcal{O}(100,000)$ simulated BBH events in this study, this is not a feasible approach.

Instead, we utilize cWB in combination with a supervised machine learning (ML) algorithm XGBoost [94] to rapidly estimate the source frame total mass for a given event. This estimate, which we call M_{cWB} , serves as our test statistic. The injected source frame total mass M_{inj} is designated as the target variable in our ML regression model, and the summary statistics reconstructed by cWB are used as input features to estimate the target variable. Figure 2 shows the injected M_{inj} and estimated M_{cWB} for simulated BBH events drawn from the Power Law + Peak mass model. In Appendix C.4 of Ref. [6], we use a simplified approach without ML, but we expect the method presented in this paper to be more sensitive to measure outliers of the stellar-mass BBH population.

To generate a robust estimate of the source frame total mass, we train our ML algorithm on a simulation set independent from the stellar-mass BBH population simulation sets. This way, the trained ML algorithm is conservative when applied to tested population models. We use the IMRPhenomPv2 waveform approximant [95] and inject simulated waveforms uniformly throughout O1, O2, and our emulated O3a observation run. For each

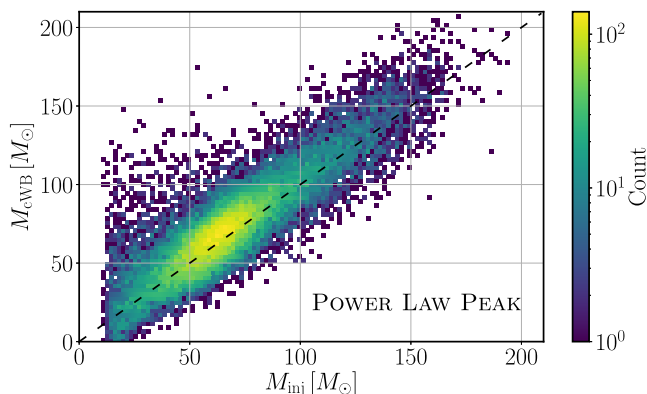


FIG. 2. Estimated source frame total mass M_{cWB} and injected source frame total mass M_{inj} for simulated BBH events drawn from the Power Law + Peak mass distribution model. Dashed black line indicates the identity line.

training event, the primary mass m_1 is drawn from a uniform distribution $m_1 \in [1, 150] M_{\odot}$, and the mass ratio q is fixed equal to one. The component spins are drawn from an isotropic distribution. Extrinsic parameters including sky location, orientation, and distance are each drawn from uniform distributions. We again consider only events detected by cWB with FAR less than 1 yr^{-1} . More information regarding our ML training and tuning procedure can be found in Appendix D.

E. Statistical procedure

Our statistical procedure is divided into two steps. First we identify which events, if any, in our sample set are potential outliers to the stellar-mass BBH population. And second, we remove any potential outliers and determine if the remaining distribution of events is consistent with the stellar-mass BBH population. We use M_{cWB} , defined in Sec. II D, as an outlier test statistic where a higher value of M_{cWB} indicates a greater probability for the event to have originated inside the PI mass gap.

First, to estimate the significance of potential outliers, we consider a bootstrap resampling procedure. In this procedure, we identify the event x which has the highest value of M_{cWB} from our sample set of N event candidates. We then randomly draw N simulated events with replacement from the model distribution and select the simulated event from the set which has the highest value of M_{cWB} , i.e., we select the outlier from the simulated set. We repeat this routine 10,000 times, and thus estimate the distribution of outliers events O predicted by the model. So, the probability for a model to produce an outlier as significant as event x is given as:

$$\Pr(M_{\text{cWB}}^x \geq M_{\text{cWB}}^O | H_0), \quad (1)$$

where H_0 represents the null hypothesis that event x is consistent with the model outlier distribution O . This is the one-sided (right tail) p-value. If this probability is less than our conservatively chosen threshold $\alpha = 0.01$, we claim that event x is a potential outlier to the tested population model.

In the case where event x is found to be a potential outlier, we omit the event from the sample set and reanalyze the remaining $N - 1$ events. We repeat this process until all outliers are identified.

Once we have identified the number of population outliers N_{out} , we then study whether the remaining sample set of $N - N_{\text{out}}$ events are consistent with the population model. We again consider a bootstrap resampling procedure where the BBH events in our sample set are ordered according to M_{cWB} . We then randomly draw 10,000 batches of $N - N_{\text{out}}$ simulated events and sort each batch according to M_{cWB} . The n th ranked event from our sample set is compared against the n th ranked simulated event from each batch. Thus, we can identify if the n th ranked event in

our sample set is consistent with predictions made by the tested population model.

To further quantify the consistency between the sample distribution and tested population model, we apply the Anderson-Darling (AD) goodness-of-fit test [96]. Compared to the Kolmogorov-Smirnov (KS) test [97,98], the AD test is expected to be more sensitive at the tails of the distribution and so is more suitable for this analysis. If the AD p-value p_{AD} is less than $\beta = 0.01$, then we find the sample distribution is inconsistent with the tested model distribution.

III. RESULTS

A. PI mass limit of $65 M_{\odot}$

In this section, we present the results of our analysis under three separate assumptions of the PI mass limit. First, we examine stellar-mass BBH models under the assumption of a $65 M_{\odot}$ PI mass limit: BPL₆₅ and PLP₆₅, defined in Sec. II B. Our sample distribution consists of 22 BBH events detected by cWB (see Table I). We order these events according to M_{cWB} . Here, the first rank event is GW190521 with $M_{cWB} = 193.2 M_{\odot}$.

We compare GW190521 against outliers predicted by each model to determine if the event is a stellar-mass BBH population outlier. Specifically, we generate batches of 22 simulated events from the BPL₆₅ and PLP₆₅ models and compare GW190521 against the outlier from each batch. The rotated histograms in Fig. 3 show M_{cWB} for

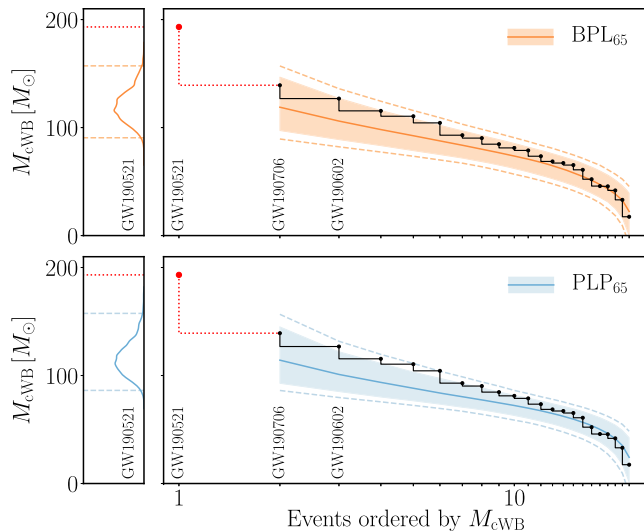


FIG. 3. Distribution of BBH events detected by cWB, shown in red/black, ranked according to M_{cWB} and compared to predictions made by stellar-mass models BPL₆₅ (top) and PLP₆₅ (bottom), which assume a PI mass limit of $65 M_{\odot}$. Event names are abbreviated. Shaded region indicates 90% confidence interval, dashed lines indicate 98% confidence interval. GW190521 is identified as an outlier to both population models. The remaining BBH events are consistent with both models.

GW190521 (red dotted line) compared against the distribution of M_{cWB} for outliers predicted by BPL₆₅ (orange) and PLP₆₅ (blue). The dashed lines in each histogram denote the 98% confidence predicted by the model. Using Eq. (1), we find the probability for the BPL₆₅ and PLP₆₅ models to produce an outlier as significant as GW190521 to be less than 0.0001. So, assuming the PI mass limit exists as a sharp cutoff at $65 M_{\odot}$, we find GW190521 to be an outlier to the stellar-mass BBH population.

Next, we investigate if any other events are outliers to the stellar-mass BBH population. We omit GW190521, and reanalyze the remaining 21 detected BBH events in our sample set against batches of 21 simulated BBH events predicted by each model. The highest ranked event in this sample set is GW190706_222641 with $M_{cWB} = 139.2 M_{\odot}$. After applying the same statistical procedure as above, we calculate a p-value of 0.109 and 0.082 for the BPL₆₅ and PLP₆₅ model, respectively, and so this event is consistent with both population models.

After identifying all potential population outliers, in this case just GW190521, we examine whether the remaining sample distribution is consistent with the tested population models. Figure 3 shows the estimated M_{cWB} for our sample set (black) compared to predictions made by the BPL₆₅ (orange) and PLP₆₅ (blue) models. The dark orange (blue) line represents the median value of M_{cWB} predicted by the BPL₆₅ (PLP₆₅) model, the shaded region indicates 90% confidence interval, and the dashed lines indicate 98% confidence interval. We find that all 21 events lie within the 98% confidence interval predicted by each model. Applying the AD goodness-of-fit test, we estimate p_{AD} for the BPL₆₅ model and PLP₆₅ model to be greater than 0.25. So, the distribution of detected events, excluding GW190521, is consistent with both stellar-mass BBH models.

B. PI mass limit of $50 M_{\odot}$

Next, we repeat our statistical analysis for stellar-mass BBH models assuming a PI mass limit of $50 M_{\odot}$: BPL₅₀ and PLP₅₀. Here, we again compare the M_{cWB} distribution of the 22 BBH events detected cWB (our sample distribution) against the M_{cWB} distribution predicted by the two models.

As in Sec. III A, we find GW190521 to be an outlier to the tested stellar-mass BBH population models (p-value < 0.0001). So, we reanalyze the remaining 21 BBH detections. Here, the highest ranked event from this set, GW190706_222641, has an outlier p-value of 0.009 when compared to the BPL₅₀ model and an outlier p-value of 0.011 when compared to the PLP₅₀ model. By definition of our *a priori* p-value threshold $\alpha = 0.01$, we treat this event as a potential outlier to the BPL₅₀ model but not the PLP₅₀ model. No other potential outliers are identified for the BPL₅₀ stellar-mass model. In Fig. 4, the rotated histograms show the estimated M_{cWB} of potential outliers compared to predictions made by each population model. Here, it is

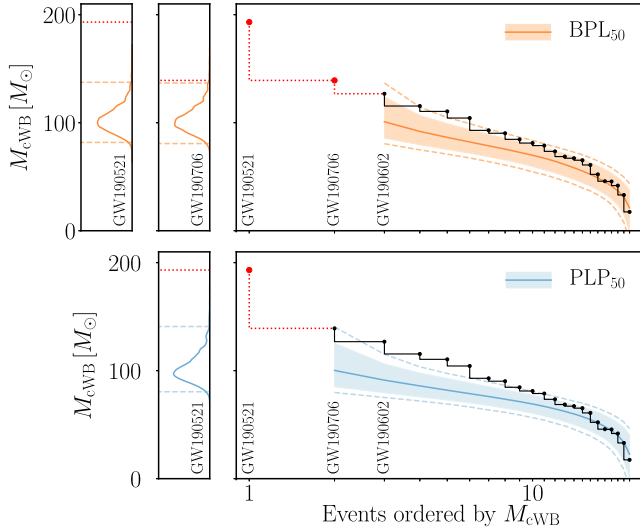


FIG. 4. Distribution of BBH events detected by cWB, shown in red/black, compared to predictions made by stellar-mass models BPL₅₀ (top) and PLP₅₀ (bottom), which assume a PI mass limit of 50 M_{\odot} . For the BPL₅₀ model, GW190521 and GW190706_222641 are evaluated as potential outliers, and the remaining events are consistent with the population model. For the PLP₅₀ model, only GW190521 is identified as an outlier, while the remaining distribution of events are inconsistent with the model.

evident that GW190521 is a clear outlier to both models, whereas GW190706_222641 lies just outside the 98% confidence interval for the BPL₅₀ model.

Next, we examine the remaining distribution of detected events. The main panels in Fig. 4 show this distribution compared to predictions made by each model. Here, we observe that the sample distribution appears to deviate from the predicted distributions, particularly the PLP₅₀ model distribution. That is, there are more high mass events in the sample distribution compared to the predictions made by each model. To quantify this deviation, we apply the AD test. For the BPL₅₀ stellar-mass model, we analyze the 20 nonoutlier BBH detections and calculate p_{AD} to be 0.039—so the detected events are consistent with the model. For the PLP₅₀ model, we analyze the 21 nonoutlier BBH detections, including GW190706_222641, and calculate p_{AD} to be 0.006. This indicates that the sample distribution is inconsistent with the PLP₅₀ model. As a cross-check, we examine the case of evaluating GW190706_222641 as a potential outlier, since it barely passes our outlier p-value threshold. Applying the AD test for this case, we calculate p_{AD} to be 0.036. So, if GW190706_222641 is treated as a potential outlier, then the remaining sample distribution is consistent with PLP₅₀.

C. PI mass limit of 40 M_{\odot}

In the final case, we consider a PI mass limit of 40 M_{\odot} . Our results for this case are summarized in Fig. 5. As in the

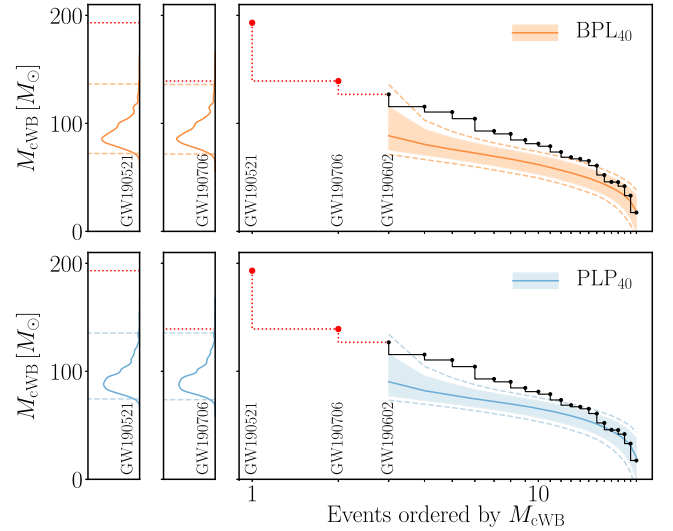


FIG. 5. Distribution of BBH events detected by cWB, shown in red/black, compared to predictions made by stellar-mass models BPL₄₀ (top) and PLP₄₀ (bottom). Two events—GW190521 and GW190706_222641—are found to be potential population outliers. The remaining 20 BBH events are inconsistent with both stellar-mass BBH population models.

previous two cases, we first compare GW190521 against the distribution of outliers predicted by each model. We again find GW190521 to be an outlier with a p-value less than 0.0001 for both models (leftmost rotated histogram). We then reanalyze the remaining 21 BBH events in our sample set to test if the highest ranked event in this set, GW190706_222641, is an outlier to the stellar-mass BBH models. For this event, we calculate a p-value of 0.007 and 0.006 for the BPL₄₀ and PLP₄₀ models, respectively (second rotated histogram). So, this event is also classified as a potential outlier. We then rerun our analysis for the remaining 20 detected BBH events and find no additional outliers in the sample set.

The main panels in Fig. 5 show the sample distribution, according to M_{cWB} , compared to predictions made by each stellar-mass BBH population model. Here, we notice that more high mass events are found compared to the predictions made by either population model. Applying the AD test, we find p_{AD} for both the BPL₄₀ model and the PLP₄₀ model to be less than 0.001. So, if the PI mass limit is at 40 M_{\odot} , then we have two potential outliers—GW190521 and GW190706_222641—and out of the remaining 20 BBH events, there are more high mass events than predicted by the models.

IV. CONCLUSIONS

In this paper, we introduce a method to determine if the distribution of BBH events detected by cWB is consistent with the stellar-mass BBH population. Detector noise is unavoidable and could affect the inference and

understanding of any given event, thereby making an otherwise ordinary BBH merger appear as an outlier. Our analysis operates directly on search statistics by injecting simulated waveforms representative of the stellar-mass BBH population directly into the detector strain data and recovering them with the search pipeline. This way, any inference errors caused by the detector noise are captured within the model distribution.

The existence of the PI mass gap is widely accepted, though its bounds are not yet well defined. Following the detections of BBH mergers, and in particular GW190521, made by LIGO-Virgo, it is now possible to probe the PI mass gap. Some works have investigated stellar evolution mechanisms which may increase the PI lower limit to account for GW190521 [28,30–32]. Alternative approaches have examined the mass models fit to the LIGO-Virgo BBH detections [5,6,75,76] to understand whether there is evidence for a mass gap [68] or whether outliers exist [70] (see also Appendix A). In this paper, we examine three separate theoretically motivated cutoffs in the black hole mass distribution to model the PI mass lower boundary to identify if any detected binaries merged within the mass gap.

We began this study to further investigate the origin of the GW190521 signal and to understand whether an event like GW190521 could be imitated by a stellar-mass BBH merger. We conclude that if the PI mass gap begins at or below $65 M_{\odot}$, then GW190521 is a clear outlier to the stellar-mass BBH population and that its outlier significance cannot be explained by statistical fluctuations. Any population study, such as the analysis presented in this paper, is dependent on the sample distribution. So, we expect the significance of stellar-mass BBH population outliers presented here to fluctuate following future observing runs. However, it is critical to note that we do not expect our outlier significance of GW190521 to change in a meaningful way (See Appendix B). Our conclusion on GW190521 is consistent with the population re-weighted posterior probability, which places the primary mass m_1 firmly above $65 M_{\odot}$ [6].

Regarding the remaining distribution of 21 BBH mergers detected by cWB, we find it unlikely that this sample distribution is consistent with the stellar-mass BBH population, assuming the PI mass limit exists at $50 M_{\odot}$. In this case, there is also some evidence that GW190706_222641 could be an outlier to the stellar-mass BBH population. And if the PI mass limit is at $40 M_{\odot}$, differences between the sample distribution and stellar-mass BBH population models cannot be reconciled. In this case, there is overwhelming evidence that cWB has detected more high mass BBH events compared to predictions made by the stellar-mass models.

In future work, we aim to expand upon this analysis by testing BBH population models corresponding to specific astrophysical channels. In this case, the inclusion of spin

effects into our test statistic will become important to distinguish between different alternative scenarios.

ACKNOWLEDGMENTS

This work was supported by NSF Grant No. PHY 1806165. We gratefully acknowledge the computational resources provided by LIGO-Virgo. This material is based upon work supported by NSFs LIGO Laboratory which is a major facility fully funded by the National Science Foundation. This research has made use of data, software and/or web tools obtained from the Gravitational Wave Open Science Center, a service of LIGO Laboratory, the LIGO Scientific Collaboration and the Virgo Collaboration. The authors thank Shanika Galaudage for providing mass samples corresponding to the GWTC-2 population models. I. B. acknowledges support by the National Science Foundation under Grant No. PHY 1911796, the Alfred P. Sloan Foundation and by the University of Florida. We thank Ik Siong Heng, Erik Katsavounidis, Peter Shawhan, and Michele Zanolin for their continued participation and effort in reviewing cWB analyses. In this analysis, we make use of open source PYTHON packages including NumPy [99], SciPy [100], Pandas [101], Matplotlib [102], Seaborn [103], and scikit-learn [104].

APPENDIX A: CONSISTENCY CHECK ON PHENOMENOLOGICAL POPULATION MODELS

This section serves as a consistency check on two phenomenological population models, Broken Power Law and Power Law + Peak, which are fit to the population of BBH mergers detected by LIGO-Virgo. This is an update to the work presented in Ref. [6], Appendix C.4.

Here, we apply the same methodology presented in Sec. II. First, we compare GW190521 against the distribution of outliers predicted by each model. We estimate the probability for the Broken Power Law model and the Power Law + Peak model to produce an outlier as significant as GW190521 to be 0.0104 and 0.015, respectively. This is consistent with the result reported in Ref. [6], where we estimate the outlier p-value of GW190521 to be 0.053 for the Broken Power Law model and 0.077 for the Power Law + Peak model. Here, we expect the change in significance is due to the improved sensitivity to population outliers with the updated method presented in this paper. The results of this test are summarized in Figure 6. Analyzing the full distribution of events detected by cWB, we find that the sample distribution is consistent with both population models ($p_{AD} > 0.25$). Overall, we cannot make any conclusive statements regarding which model fits the observed data better. Future BBH detections could help resolve the differences between the models and point to a preferred model.

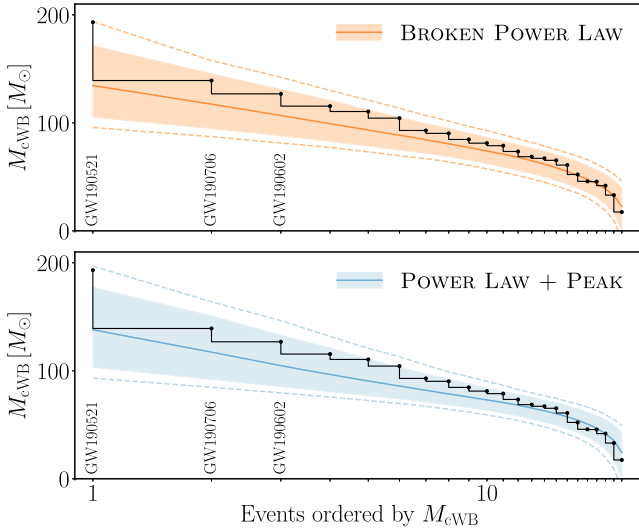


FIG. 6. Distribution of BBH events detected by cWB, shown in black, ranked according to M_{cWB} and compared to predictions made by the Broken Power Law model (top) and Power Law + Peak model (bottom). Shaded region indicates 90% confidence interval, dashed lines indicate 98% confidence interval. The BBH events detected by cWB are consistent with both population models.

APPENDIX B: FURTHER INVESTIGATION OF THE OUTLIER SIGNIFICANCE OF GW190521

One main conclusion of this work is that GW190521 is inconsistent with the stellar-mass BBH population, assuming a PI mass limit at or below $65 M_{\odot}$. This result, however, is inherently dependent on our sample size and is subject to change following additional detections in future observing runs. Nevertheless, we expect the outlier significance of GW190521 presented here will not meaningfully change following future observing runs.

To test this, we consider a hypothetical scenario where the cWB pipeline identified 100 BBH events during the O1, O2, and O3a observing runs (instead of 22 BBH events). Under this scenario, we conservatively assume that GW190521 persists as the first rank event, according to M_{cWB} . Following the statistical procedure described in Sec. II E, we estimate the outlier p-value of GW190521 to be less than 0.0001. Though this p-value is admittedly limited by our simulated sample size, our conclusion remains that no simulated BBH event representative of the stellar-mass BBH population is as significant as GW190521.

APPENDIX C: ML TRAINING FEATURE LIST

We utilize XGBoost, a boosted decision tree ML algorithm [94], as a regression tool to estimate the source frame total mass M_{cWB} for a given event. To estimate the target variable, ML algorithms require input features which are correlated with the target variable. In our case, the target variable is the source frame total mass M_{inj} , and the ML

features are summary statistics reconstructed by cWB. Below is the list of 13 summary statistics used in the training process, ordered according to the *gain* importance metric. Unsurprisingly, the most important feature is a measure of the central frequency f_0 , which is expected to be inversely proportional to the detector frame total mass. Further attempts at feature pruning reduced our evaluation score on a validation data set. Some summary statistic definitions below are equivalent to those found in Ref. [105].

- (i) f_0 —Energy weighted signal central frequency.
- (ii) ΔF_s —Energy weighted signal bandwidth.
- (iii) O_n —Observation run number. Including the observation run number allows the ML algorithm to respond to changes in detector sensitivity across separate observing runs.
- (iv) n_f —Effective number of time-frequency resolutions used for event detection and waveform reconstruction.
- (v) ΔT_s —Energy weighted signal duration.
- (vi) F_M —Chirp mass energy fraction, chirp mass goodness of fit metric, defined in Ref. [106].
- (vii) \mathcal{M} —Chirp mass parameter estimated in the time-frequency domain, defined in Ref. [106].
- (viii) SNR_{net} —Quadrature sum of the reconstructed signal SNR found in each GW detector.
- (ix) Q_0 —An estimation of the effective number of cycles in a cWB event [107].
- (x) e_M —Chirp mass ellipticity, chirp mass goodness of fit metric, defined in Ref. [106].
- (xi) c_c —Coherent energy divided by the sum of coherent energy and null energy, defined in Ref. [71].
- (xii) Q_1 —The waveform shape parameter [107] developed to identify a characteristic family of (blip) glitches present in the detectors [3,108].
- (xiii) E_c/L —Ratio of the coherent energy to the network likelihood.

TABLE II. List of tuned hyper-parameters. Bold parameters indicate hyper-parameters selected through tuning with 5-fold cross validation. The `base_score` is initialized to the mean source frame total mass of the training data set \bar{M}_{inj} †: the `n_estimators` hyperparameter is optimized using early stopping, described in the main text.

Hyperparameter	Entry
<code>objective</code>	<code>reg:pseudohubererror</code>
<code>tree_method</code>	<code>gpu_hist</code>
<code>grow_policy</code>	<code>lossguide</code>
<code>base_score</code>	\bar{M}_{inj}
<code>n_estimators</code>	20000†
<code>learning_rate</code>	0.03 , 0.1
<code>max_depth</code>	5 , 7, 9, 11, 13
<code>min_child_weight</code>	1, 5 , 10
<code>gamma</code>	1.0 , 2.0, 5.0
<code>colsample_bytree</code>	0.6, 0.8, 1.0
<code>subsample</code>	0.6, 0.8

APPENDIX D: ML TRAINING AND TUNING

Our training set consists of simulated BBH events recovered by cWB, described in Sec. II D. We use the Huber loss function [109] as the learning objective for our XGBoost regression model. This loss function is expected to be less sensitive to outliers compared to the squared loss function. The `base_score`, which represents the initial prediction for all instances, is initialized to the mean injected source frame total mass of the training data set

\bar{M}_{inj} . To optimize the number of decision trees, we employ early stopping which restricts the number of trees based on the model performance over an independent validation data set. Here, early stopping effectively limits overfitting. To tune the remaining XGBoost hyperparameters, we perform a grid search over 540 combinations of hyperparameters and use 5-fold cross validation to further prevent overfitting. Selected hyperparameters are shown in bold in Table II.

-
- [1] J. Aasi *et al.*, *Classical Quantum Gravity* **32**, 074001 (2015).
- [2] F. Acernese, M. Agathos, and K. Agatsuma, *Classical Quantum Gravity* **32**, 024001 (2015).
- [3] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. X* **9**, 031040 (2019).
- [4] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. X* **11**, 021053 (2021).
- [5] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J.* **882**, L24 (2019).
- [6] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J. Lett.* **913**, L7 (2021).
- [7] S. Woosley, S. Blinnikov, and A. Heger, *Nature (London)* **450**, 390 (2007).
- [8] M. Spera and M. Mapelli, *Mon. Not. R. Astron. Soc.* **470**, 4739 (2017).
- [9] S. E. Woosley, *Astrophys. J.* **836**, 244 (2017).
- [10] N. Giacobbo, M. Mapelli, and M. Spera, *Mon. Not. R. Astron. Soc.* **474**, 2959 (2018).
- [11] K. Belczynski *et al.*, *Astron. Astrophys.* **594**, A97 (2016).
- [12] Z. Barkat, G. Rakavy, and N. Sack, *Phys. Rev. Lett.* **18**, 379 (1967).
- [13] K.-J. Chen, S. Woosley, A. Heger, A. Almgren, and D. J. Whalen, *Astrophys. J.* **792**, 28 (2014).
- [14] T. Yoshida, H. Umeda, K. Maeda, and T. Ishii, *Mon. Not. R. Astron. Soc.* **457**, 351 (2016).
- [15] W. A. Fowler and F. Hoyle, *Astrophys. J. Suppl. Ser.* **9**, 201 (1964).
- [16] A. Heger, C. L. Fryer, S. E. Woosley, N. Langer, and D. H. Hartmann, *Astrophys. J.* **591**, 288 (2003).
- [17] J. R. Bond, W. D. Arnett, and B. J. Carr, *Astrophys. J.* **280**, 825 (1984).
- [18] W. W. Ober, M. F. El Eid, and K. J. Fricke, *Astron. Astrophys.* **119**, 61 (1983), <https://ui.adsabs.harvard.edu/abs/1983A&A...119...61O>.
- [19] R. P. Kudritzki, A. Pauldrach, and J. Puls, *Astron. Astrophys.* **173**, 293 (1987), <https://ui.adsabs.harvard.edu/abs/1987A&A...173..293K>.
- [20] Jorick S. Vink, A. de Koter, and H. J. G. L. M. Lamers, *Astron. Astrophys.* **369**, 574 (2001).
- [21] M. Mapelli, M. Spera, E. Montanari, M. Limongi, A. Chieffi, N. Giacobbo, A. Bressan, and Y. Bouffanais, *Astrophys. J.* **888**, 76 (2020).
- [22] M. Mapelli, M. Colpi, and L. Zampieri, *Mon. Not. R. Astron. Soc.: Lett.* **395**, L71 (2009).
- [23] M. Mapelli, E. Ripamonti, L. Zampieri, M. Colpi, and A. Bressan, *Mon. Not. R. Astron. Soc.* **408**, 234 (2010).
- [24] M. Mapelli, L. Zampieri, E. Ripamonti, and A. Bressan, *Mon. Not. R. Astron. Soc.* **429**, 2298 (2013).
- [25] K. Belczynski, T. Bulik, C. L. Fryer, A. Ruitter, F. Valsecchi, J. S. Vink, and J. R. Hurley, *Astrophys. J.* **714**, 1217 (2010).
- [26] C. L. Fryer, K. Belczynski, G. Wiktorowicz, M. Dominik, V. Kalogera, and D. E. Holz, *Astrophys. J.* **749**, 91 (2012).
- [27] M. Spera, M. Mapelli, and A. Bressan, *Mon. Not. R. Astron. Soc.* **451**, 4086 (2015).
- [28] R. Farmer, M. Renzo, S. E. de Mink, M. Fishbach, and S. Justham, *Astrophys. J.* **902**, L36 (2020).
- [29] M. Limongi and A. Chieffi, *Astrophys. J. Suppl. Ser.* **237**, 13 (2018).
- [30] G. Costa, A. Bressan, M. Mapelli, P. Marigo, G. Iorio, and M. Spera, *Mon. Not. R. Astron. Soc.* **501**, 4514 (2020).
- [31] K. Belczynski, *Astrophys. J. Lett.* **905**, L15 (2020).
- [32] S. E. Woosley and A. Heger, *Astrophys. J. Lett.* **912**, L31 (2021).
- [33] S. E. Woosley, *Astrophys. J.* **878**, 49 (2019).
- [34] R. Abbott *et al.*, *Astrophys. J.* **900**, L13 (2020).
- [35] M. Coleman Miller and D. P. Hamilton, *Mon. Not. R. Astron. Soc.* **330**, 232 (2002).
- [36] R. M. O’Leary, F. A. Rasio, J. M. Fregeau, N. Ivanova, and R. O’Shaughnessy, *Astrophys. J.* **637**, 937 (2006).
- [37] M. Giersz, N. Leigh, A. Hypki, N. Litzgendorf, and A. Askar, *Mon. Not. R. Astron. Soc.* **454**, 3150 (2015).
- [38] M. Mapelli, *Mon. Not. R. Astron. Soc.* **459**, 3432 (2016).
- [39] M. Spera, M. Mapelli, N. Giacobbo, A. A. Trani, A. Bressan, and G. Costa, *Mon. Not. R. Astron. Soc.* **485**, 889 (2019).
- [40] U. N. Di Carlo, N. Giacobbo, M. Mapelli, M. Pasquato, M. Spera, L. Wang, and F. Haardt, *Mon. Not. R. Astron. Soc.* **487**, 2947 (2019).
- [41] Y. Bouffanais, M. Mapelli, D. Gerosa, U. N. D. Carlo, N. Giacobbo, E. Berti, and V. Baibhav, *Astrophys. J.* **886**, 25 (2019).
- [42] U. N. Di Carlo, M. Mapelli, Y. Bouffanais, N. Giacobbo, F. Santoliquido, A. Bressan, M. Spera, and F. Haardt, *Mon. Not. R. Astron. Soc.* **497**, 1043 (2020).

- [43] V. Gayathri, I. Bartos, Z. Haiman, S. Klimentko, B. Kocsis, S. Márka, and Y. Yang, *Astrophys. J. Lett.* **890**, L20 (2020).
- [44] D. Veske, Z. Márka, A. G. Sullivan, I. Bartos, K. Rainer Corley, J. Samsing, and S. Márka, *Mon. Not. R. Astron. Soc.* **498**, L46 (2020).
- [45] Y. Yang, I. Bartos, V. Gayathri, K. Ford, Z. Haiman, S. Klimentko, B. Kocsis, S. Mrka, Z. Mrka, B. McKernan *et al.*, *Phys. Rev. Lett.* **123**, 181101 (2019).
- [46] C. Kimball *et al.*, *Astrophys. J. Lett.* **915**, L35 (2021).
- [47] G. Fragione, B. Kocsis, F. A. Rasio, and J. Silk, [arXiv:2107.04639](https://arxiv.org/abs/2107.04639).
- [48] G. Fragione and J. Silk, *Mon. Not. R. Astron. Soc.* **498**, 4591 (2020).
- [49] I. Bartos, B. Kocsis, Z. Haiman, and S. Márka, *Astrophys. J.* **835**, 165 (2017).
- [50] B. McKernan, K. E. S. Ford, J. Bellovary, N. W. C. Leigh, Z. Haiman, B. Kocsis, W. Lyra, M.-M. M. Low, B. Metzger, M. O'Dowd, S. Endlich, and D. J. Rosen, *Astrophys. J.* **866**, 66 (2018).
- [51] B. McKernan, K. E. S. Ford, W. Lyra, and H. B. Perets, *Mon. Not. R. Astron. Soc.* **425**, 460 (2012).
- [52] B. McKernan, K. E. S. Ford, B. Kocsis, W. Lyra, and L. M. Winter, *Mon. Not. R. Astron. Soc.* **441**, 900 (2014).
- [53] J. M. Bellovary, M.-M. M. Low, B. McKernan, and K. E. S. Ford, *Astrophys. J.* **819**, L17 (2016).
- [54] B. McKernan, K. E. S. Ford, R. OShaughnessy, and D. Wysocki, *Mon. Not. R. Astron. Soc.* **494**, 1203 (2020).
- [55] Y. Yang, I. Bartos, Z. Haiman, B. Kocsis, Z. Márka, N. C. Stone, and S. Márka, *Astrophys. J.* **876**, 122 (2019).
- [56] Y. Yang, V. Gayathri, I. Bartos, Z. Haiman, M. Safarzadeh, and H. Tagawa, *Astrophys. J.* **901**, L34 (2020).
- [57] H. Tagawa, B. Kocsis, Z. Haiman, I. Bartos, K. Omukai, and J. Samsing, *Astrophys. J.* **908**, 194 (2021).
- [58] S. F. Portegies Zwart, H. Baumgardt, P. Hut, J. Makino, and S. L. W. McMillan, *Nature (London)* **428**, 724 (2004).
- [59] K. Kremer, M. Spera, D. Becker, S. Chatterjee, U. N. Di Carlo, G. Fragione, C. L. Rodriguez, C. S. Ye, and F. A. Rasio, *Astrophys. J.* **903**, 45 (2020).
- [60] E. González, K. Kremer, S. Chatterjee, G. Fragione, C. L. Rodriguez, N. C. Weatherford, C. S. Ye, and F. A. Rasio, *Astrophys. J. Lett.* **908**, L29 (2021).
- [61] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **125**, 101102 (2020).
- [62] V. Gayathri, J. Healy, J. Lange, B. O'Brien, M. Szczepanczyk, I. Bartos, M. Campanelli, S. Klimentko, C. Lousto, and R. O'Shaughnessy, [arXiv:2009.05461](https://arxiv.org/abs/2009.05461).
- [63] V. D. Luca, V. Desjacques, G. Franciolini, P. Pani, and A. Riotto, *Phys. Rev. Lett.* **126**, 051101 (2021).
- [64] M. Fishbach and D. E. Holz, *Astrophys. J.* **904**, L26 (2020).
- [65] M. Safarzadeh and Z. Haiman, *Astrophys. J.* **903**, L21 (2020).
- [66] I. Romero-Shaw, P. D. Lasky, E. Thrane, and J. C. Bustillo, *Astrophys. J.* **903**, L5 (2020).
- [67] E. J. Baxter, D. Croon, S. D. McDermott, and J. Sakstein, *Astrophys. J. Lett.* **916**, L16 (2021).
- [68] B. Edelman, Z. Doctor, and B. Farr, *Astrophys. J. Lett.* **913**, L23 (2021).
- [69] G. Fragione, A. Loeb, and F. A. Rasio, *Astrophys. J. Lett.* **902**, L26 (2020).
- [70] M. Fishbach, W. M. Farr, and D. E. Holz, *Astrophys. J.* **891**, L31 (2020).
- [71] S. Klimentko, I. Yakushin, A. Mercer, and G. Mitselmakher, *Classical Quantum Gravity* **25**, 114029 (2008).
- [72] S. Klimentko, G. Vedovato, M. Drago, F. Salemi, V. Tiwari, G. A. Prodi, C. Lazzaro, K. Ackley, S. Tiwari, C. F. Da Silva, and G. Mitselmakher, *Phys. Rev. D* **93**, 042004 (2016).
- [73] V. Necula, S. Klimentko, and G. Mitselmakher, *J. Phys. Conf. Ser.* **363**, 012032 (2012).
- [74] M. Szczepańczyk *et al.*, *Phys. Rev. D* **103**, 082002 (2021).
- [75] E. D. Kovetz, I. Cholis, P. C. Breyse, and M. Kamionkowski, *Phys. Rev. D* **95**, 103010 (2017).
- [76] C. Talbot and E. Thrane, *Astrophys. J.* **856**, 173 (2018).
- [77] E. Thrane and C. Talbot, *Pub. Astron. Soc. Aust.* **36**, e010 (2019).
- [78] I. Mandel, W. M. Farr, and J. R. Gair, *Mon. Not. R. Astron. Soc.* **486**, 1086 (2019).
- [79] S. Vitale, D. Gerosa, W. M. Farr, and S. R. Taylor, [arXiv:2007.05579](https://arxiv.org/abs/2007.05579).
- [80] E. E. Salpeter, *Astrophys. J.* **121**, 161 (1955).
- [81] P. Kroupa, *Mon. Not. R. Astron. Soc.* **322**, 231 (2001).
- [82] K. Belczynski *et al.*, *Astron. Astrophys.* **636**, A104 (2020).
- [83] A. Heger and S. E. Woosley, *Astrophys. J.* **567**, 532 (2002).
- [84] S. E. Woosley and A. Heger, The deaths of very massive stars, in *Very Massive Stars in the Local Universe*, edited by J. S. Vink (Springer International Publishing, Cham, 2015), Chap. 7, pp. 199–225.
- [85] A. Boh, L. Shao, A. Taracchini, A. Buonanno, S. Babak, I. W. Harry, I. Hinder, S. Ossokine, M. Prer, V. Raymond *et al.*, *Phys. Rev. D* **95**, 044028 (2017).
- [86] R. Abbott *et al.*, *SoftwareX* **13**, 100658 (2021).
- [87] V. Kalogera, *Astrophys. J.* **541**, 319 (2000).
- [88] I. Mandel and R. O'Shaughnessy, *Classical Quantum Gravity* **27**, 114007 (2010).
- [89] M. Dominik, K. Belczynski, C. Fryer, D. E. Holz, E. Berti, T. Bulik, I. Mandel, and R. O'Shaughnessy, *Astrophys. J.* **779**, 72 (2013).
- [90] N. Giacobbo, M. Mapelli, and M. Spera, *Mon. Not. R. Astron. Soc.* **474**, 2959 (2018).
- [91] J. J. Eldridge, E. R. Stanway, L. Xiao, L. A. S. McClelland, G. Taylor, M. Ng, S. M. L. Greis, and J. C. Bray, *Pub. Astron. Soc. Aust.* **34**, e058 (2017).
- [92] A. Olejak, M. Fishbach, K. Belczynski, D. E. Holz, J. P. Lasota, M. C. Miller, and T. Bulik, *Astrophys. J. Lett.* **901**, L39 (2020).
- [93] P. A. R. Ade *et al.*, *Astron. Astrophys.* **594**, A13 (2016).
- [94] T. Chen and C. Guestrin, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16 (ACM, New York, NY, USA, 2016), pp. 785–794.
- [95] S. Khan, K. Chatzioannou, M. Hannam, and F. Ohme, *Phys. Rev. D* **100**, 024059 (2019).
- [96] T. W. Anderson and D. A. Darling, *Ann. Math. Stat.* **23**, 193 (1952).
- [97] A. N. Kolmogorov, *Giornale dell'Istituto Italiano degli Attuari* **4**, 83 (1933).

- [98] N. Smirnov, *Ann. Math. Stat.* **19**, 279 (1948).
- [99] C. R. Harris *et al.*, *Nature (London)* **585**, 357 (2020).
- [100] P. Virtanen *et al.*, *Nat. Methods* **17**, 261 (2020).
- [101] pandas-dev/pandas: Pandas,” (2020).
- [102] J. D. Hunter, *Comput. Sci. Eng.* **9**, 90 (2007).
- [103] M. L. Waskom, *J. Open Source Softw.* **6**, 3021 (2021).
- [104] F. Pedregosa *et al.*, *J. Mach. Learn. Res.* **12**, 2825 (2011), <https://ui.adsabs.harvard.edu/abs/2012arXiv1201.0490P>.
- [105] T. Mishra, B. O’Brien, V. Gayathri, M. Szczepanczyk, S. Bhaumik, I. Bartos, and S. Klimenko, *Phys. Rev. D* **104**, 023014 (2021).
- [106] V. Tiwari, S. Klimenko, V. Neucula, and G. Mitselmakher, *Classical Quantum Gravity* **33**, 01LT01 (2016).
- [107] G. Vedovato, The Qveto algorithm (2018).
- [108] J. McIver, *Classical Quantum Gravity* **29**, 124010 (2012).
- [109] P. J. Huber, *Ann. Math. Stat.* **35**, 73 (1964).