

New binary black hole mergers in the second observing run of Advanced LIGO and Advanced Virgo

Tejaswi Venumadhav,^{1,*} Barak Zackay¹,,¹ Javier Roulet²,,² Liang Dai¹,,¹ and Matias Zaldarriaga¹

¹*School of Natural Sciences, Institute for Advanced Study,
1 Einstein Drive, Princeton, New Jersey 08540, USA*

²*Department of Physics, Princeton University, Princeton, New Jersey 08540, USA*



(Received 20 August 2019; accepted 8 April 2020; published 27 April 2020)

We report the detection of new binary black hole merger events in the publicly available data from the second observing run of Advanced LIGO and Advanced Virgo (O2). The mergers were discovered using the new search pipeline described in Venumadhav *et al.* [*Phys. Rev. D* **100**, 023011 (2019)] and are above the detection thresholds as defined in Abbott *et al.* (LIGO Scientific and Virgo Collaborations) [*Phys. Rev. X* **9**, 031040 (2019)]. Three of the mergers (GW170121, GW170304, GW170727) have inferred probabilities of being of astrophysical origin $p_{\text{astro}} > 0.98$. The remaining three (GW170425, GW170202, GW170403) are less certain, with p_{astro} ranging from 0.5 to 0.8. The newly found mergers largely share the statistical properties of previously reported events, with the exception of GW170403, the least secure event, which has a highly negative effective spin parameter χ_{eff} . The most secure new event, GW170121 ($p_{\text{astro}} > 0.99$), is also notable due to its inferred negative value of χ_{eff} , which is inconsistent with being positive at the $\approx 95.8\%$ confidence level. The new mergers nearly double the sample of gravitational wave events reported from O2 and present a substantial opportunity to explore the statistics of the binary black hole population in the Universe. The number of detected events is not surprising since we estimate that the detection volume of our pipeline may be larger than that of other pipelines by as much as a factor of 2 (with significant uncertainties in the estimate). The increase in volume is larger when the constituent detectors of the network have very different sensitivities, as is likely to be the case in current and future runs.

DOI: [10.1103/PhysRevD.101.083030](https://doi.org/10.1103/PhysRevD.101.083030)

I. INTRODUCTION

The LIGO-Virgo Collaboration (LVC) reported the detection of gravitational waves from ten binary black hole (BBH) and one binary neutron star mergers in their two latest observing runs, O1 and O2 [1–8]. The intrinsic properties of the mergers, namely the masses, the mass ratio, and the spins of the black holes, are important observables that can inform us about how and where the binaries were assembled. All the BBH events are consistent with mergers of black holes with comparable masses (notably, GW170729 is mildly inconsistent with equal component masses [9]); in two of the events, at least one of the components had a nonzero spin.

Currently, inference about the origin of the BBHs is limited by the small number of detected events. Future observations with improved sensitivity will enlarge the sample and map out the parameter space of the BBH population; meanwhile, it is important to adopt analysis techniques that maximize the yield of existing data.

A number of data analysis pipelines have been developed to search for transient events in LIGO-Virgo data. The two modeled searches used by the LVC are PyCBC [10] and GstLAL [11]; these pipelines use matched filtering with a template bank of target compact binary coalescence signals. In addition, the LVC runs an unmodeled transient search with the coherent wave burst pipeline [12]. There are also groups external to the LVC running independent pipelines [13,14] on the public data released by the LVC, which now includes the O1 and O2 observing runs [15,16]. In Ref. [17], henceforth TV19, we presented a new and independent pipeline to analyze public LIGO data, which we applied to the public data from the O1 observing run. The cumulative improvements significantly increased the sensitive volume (at the same detection thresholds as those of Ref. [1]), and led to the detection of a new event, which is consistent with the merger of rapidly spinning and heavy black holes [18]. In this paper, we present results from our search of coincident triggers in the Hanford (H1) and Livingston (L1) public data from the second observing run of advanced LIGO (O2) [15,16].

The paper is organized as follows: in Sec. II we begin with a preview of our final results to motivate what follows.

*tejaswi@ias.edu

We then discuss the details of the changes in the pipeline used in this paper compared to the one used to analyze the O1 data in Sec. III. Section IV summarizes the results of our analysis on the events that were previously detected by the LVC analysis pipelines in Ref. [1]. We describe the new events we found in Sec. V, estimate our improvement in sensitivity in Sec. VI, and conclude with some remarks in Sec. VII. In Appendixes A–C, we present the posterior distributions for the parameters of the new events, as well as some technical details.

II. PREVIEW OF RESULTS

We improved upon the methods we previously developed to analyze the O1 data (see TV19 and Sec. III for details), and we applied them to look for massive binary black hole mergers in the H1 and L1 data from the O2 data released by the LVC. We recover with high confidence all the BBH events that the LVC searches previously found in the bulk data release, with the notable exception of GW170818, which is one of our highest ranked candidates, but is effectively close to a single-detector trigger that would be hard to confirm using the methods adopted in this search (see Sec. IV for details).

This serves as an independent validation of the events presented in GWTC-1 [1], and a confirmation of the sensitivity of our pipeline. The fact that we can only place upper bounds on the false alarm rates (FARs) for these events, in particular, even for the least significant event in GWTC-1, GW170729, is consistent with an improvement

in sensitivity over the range of parameters included in our search. Also consistent with this is the fact that we find six new events with (a) false alarm rates below the threshold of 1 in 30 days defined in GWTC-1 (even after applying appropriate trials factors), and (b) probabilities of astrophysical origin $p_{\text{astro}} > 0.5$, as effectively quantified by their consistency with the rates of occurrence of louder events.

Figure 1 places these new events in context of the population of events in the GWTC-1 catalog by showing their marginalized posterior probability distributions in the plane of the effective spin parameter, $\chi_{\text{eff}} = (m_1\chi_{1,z} + m_2\chi_{2,z})/(m_1 + m_2)$ (where m_i and $\chi_{i,z}$, respectively, are the masses and projections of the dimensionless spin onto the orbital angular momentum), and the total source-frame mass. We chose to present the points for the new events colored according to their estimated values of p_{astro} to emphasize that these events come as a population, with these numbers being an integral part of their interpretation. In the rest of the paper, we will present the changes to our analysis methods, the procedure we use to assign values of FAR and p_{astro} to events, and a deeper view into the results.

III. CHANGES TO THE O1 ANALYSIS PIPELINE

Our analysis pipeline is similar in overall structure to the one we used in the O1 analysis. The overall flow of the search is described in Sec. II of TV19: for completeness, we briefly repeat the description here. We divide our search space into a set of banks, and within each bank, we

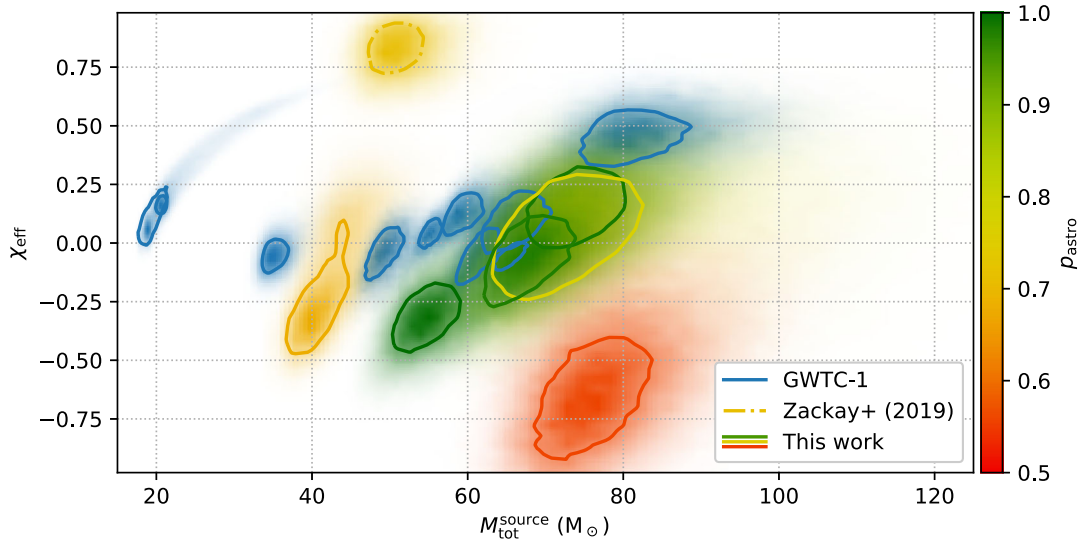


FIG. 1. Source-frame total mass and effective spin for the BBH events found in Hanford-Livingston coincidence, over O1 and O2. We recovered all the previously reported events that were in the data we searched in with high confidence, i.e., the probability of astrophysical origin $p_{\text{astro}} \approx 1$, except for GW170818; see Sec. IV. We found seven additional events ranging from marginal triggers to confident detections: one in O1 [18] and six in O2 (this work); we color code the posteriors according to the value of p_{astro} . The densities use samples from the posterior distribution of each event, and the curves are 1σ contours (i.e., they enclose $1 - e^{-1/2} \approx 39\%$ of the posterior probability). We used a prior that is uniform in detector frame m_1 , m_2 , χ_{eff} , and luminosity volume. Appendix A shows other projections of the posteriors.

construct a set of templates that approximate the waveforms generated using the IMRPhenomD approximant [19] to a desired degree of fidelity in the presence of characteristic detector noise. We then generate a time series of matched-filtering scores for each template against the data streams from the Hanford and Livingston detectors. When these scores cross a threshold value in a detector, we call this a *trigger* in that detector: we collect coincident triggers between H1 and L1 (with templates indexed by the same coefficients, and within 10 ms of each other). Subsequently we locally refine the triggers in each detector onto a finer template bank, apply vetoes based on signal quality, and pick the best coincident trigger from the subsets of refined H1 and L1 triggers. Finally we assign triggers (and hence the underlying signal/background event) to banks, compute a ranking statistic for the triggers in a bank, and assign a false alarm rate to triggers according to the empirically measured distribution of the ranking statistic.

The interested reader can find a full description of the details of the methods in TV19. In this section, we focus on the differences between the current and previous analyses. In detail, the pipeline for the O2 analysis differs in the following aspects:

- (1) *Construction of the template bank*: Our search covers the parameter space of compact binary mergers with component masses between 3 and $100 M_{\odot}$, with mass ratios $q \in [1/18, 1]$ and the projections of the dimensionless spins of the constituents on the orbital angular momentum satisfying $|\chi_i| < 0.99$. As in the O1 analysis, we partition this search space into five banks (BBH 0–4) based on the detector-frame chirp masses, and divide each bank into subbanks based on the shape of the frequency-domain amplitude profile. Within each subbank, we use the procedure described in Ref. [20] to construct grids of templates. We now use the k -means algorithm to automatically divide each bank into subbanks, each of which has a frequency-domain amplitude, $\bar{A}(f)$, that is the root-mean-square average of the amplitudes of the constituent astrophysical waveforms. In Fig. 2 we show $\bar{A}(f)$ for all the subbanks in our banks covering detector chirp masses $20\text{--}40 M_{\odot}$ (bank BBH 3) and above $40 M_{\odot}$ (BBH 4). We show the profiles only for these banks to avoid overcrowding the figure, and since all the events we report in this paper fall in these banks. The curves for the other banks (BBH 2, BBH 1, and BBH 0) continue the trend that is visible in Fig. 2, in which the cutoff frequency shifts rightward as we go to lower masses.

The phases of the templates are linear combinations of basis functions, $\psi_{\alpha}(f)$, whose form depends on the noise power spectral density (PSD); we use the PSD estimated from a representative set of files from the run (instead of a model PSD as was done in

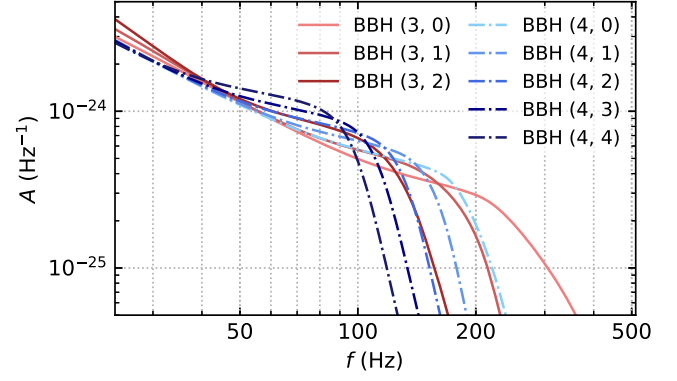


FIG. 2. Subbank amplitude profiles as a function of frequency f for banks BBH 3 and BBH 4.

the O1 analysis). We also restrict the templates to frequencies between 24 and 512 Hz. Cumulatively, these changes enable us to cover the same astrophysical parameter space using $\sim 30\%$ fewer templates, and achieve slightly better effectualness [20].

- (2) *Preprocessing and flagging the data*: Our analysis pipeline produces a stream of whitened data per 4096 s file. As in our O1 analysis, we perform several tests on the data to identify prolonged or transient disturbances in the detector that appear as segments with excess power, and discard these segments to avoid polluting our search. The overall nature and number of these tests is unchanged from the O1 analysis; a full description of these tests can be found in Sec. III C of TV19. What is different, however, is how we set thresholds to trigger these tests. If the thresholds are too low, they can be triggered by real astrophysical events. Conversely, if they are too high, the tests are not easily triggered by bad data segments. To mitigate the former effect, in the O1 analysis, we set the threshold for each test to the power achieved by signals with a fiducial signal-to-noise ratio $(\text{SNR})_{\text{max}}$ in the absence of noise. This method of setting thresholds is problematic, since in the presence of additive Gaussian random noise, real signals with $\text{SNR} < (\text{SNR})_{\text{max}}$ can have excess power above these thresholds due to upward fluctuations. Hence, in our O1 analysis we used a relatively high value of $(\text{SNR})_{\text{max}} = 30$ to ensure completeness at lower values [14].

In the analysis described in this paper, we instead set the thresholds for tests by demanding a given false-positive probability for signals with a desired signal-to-noise ratio. For whitened data, power in a given band, and on a given timescale, is distributed according to a noncentral chi-squared distribution in the presence of a signal; hence, we can set thresholds such that signals with a given value of $(\text{SNR})_{\text{max}}$ are flagged with a probability $< 10^{-4}$. Given this

promise, we are able to set thresholds with a lower target value $(\text{SNR})_{\text{max}} = 20$, without losing relevant signals in this search. By nature, the thresholds resulting from this criterion, and hence the amount of data flagged as bad, are dependent on the bank (i.e., at the same value of SNR, the thresholds on excess power are higher for a signal in BBH 4 compared to those in BBH 0). Regardless of bank, the total length of bad data never exceeds 2% of the overall runtime of O2, and hence, the effective reduction in “runtime” is not an important factor to take into account in spacetime volume estimates for the search.

- (3) *Refining coincident triggers:* The phases of our templates belong to a vector space, $\mathcal{V}$, spanned by the basis functions $\{\psi_\alpha(f) : \alpha = 1, 2, \dots, n\}$; the template bank is a discrete subset of this space with basis coefficients that live on an n -dimensional grid. We determine the extent of the grid in all dimensions by projecting a large random sample of astrophysical waveforms into $\mathcal{V}$ and ensuring that every waveform has a nearby grid point. We allow comparatively large mismatches ($\lesssim 10\%$) between astrophysical waveforms and the best template in the bank, which enables us to work with coarser grids on $\mathcal{V}$ when generating triggers. We then reduce the mismatch for significant triggers by refining their coefficients on a finer local grid. In the O1 analysis, we chose this grid to be a uniform regular grid centered on a trigger of interest [14].

For heavy BBH waveforms, the set $\mathcal{S}$ of projected astrophysical waveforms is typically thin and mildly curved in higher dimensions ($\alpha \gtrsim 2$) [20]. The strategy used in the O1 analysis can cause us to step outside $\mathcal{S}$ in these dimensions and introduce unphysical degrees of freedom that pick up noise but no signal. In this analysis, we change the spacing of the finer grid and excise unnecessary elements to ensure that we enumerate over templates within $\mathcal{S}$. Note that different choices of how this refinement is done can lead to different quoted SNRs for the same astrophysical signal depending on how closely the finer grid approaches it. Hence, we apply the same strategy to the background triggers (found via time slides) to avoid biasing the calculation of false-alarm rates.

- (4) *Reducing cross-contamination between banks:* High SNR triggers tend to appear in several of our chirp-mass banks, both in the time slides used to estimate our background and in the set of coincident triggers. In the O1 analysis, we assigned triggers to the best subbank in a given chirp-mass bank (as determined by the incoherent network $\text{SNR}^2 = \rho_H^2 + \rho_L^2$, where ρ_H^2 and ρ_L^2 are the incoherent squared SNRs in Hanford and Livingston, respectively), but allowed

them to appear in multiple banks [17]. This choice was conservative, in that it caused us to overestimate the FAR for real events. In this analysis, we improve upon this in two ways: we assign both background and coincident triggers to a unique bank (and subbank within), and instead of the incoherent network SNR^2 , we use a discriminator that better accounts for the different structures of the subbanks. Appendix B contains the expression [see Eq. (B7)] and outlines a derivation.

- (5) *Computing the false-alarm rate:* After collecting triggers and assigning them to banks, we estimate the FAR for a given coincident trigger by comparing it to the background triggers (collected using time slides) within the same chirp-mass bank. Our statistic for comparing triggers is the coherent score [21], which for a given trigger t is an approximation of the likelihood ratio under the signal $[\mathcal{L}(t|\mathcal{S})]$ and noise $[\mathcal{L}(t|\mathcal{N})]$ hypotheses. To estimate $\mathcal{L}(t|\mathcal{N})$, we approximate the distribution of background triggers in each detector (as a function of SNR^2) by its survival function. The survival function (or more correctly, the complementary cumulative distribution function) has the virtue that we can estimate it empirically by just ranking triggers as a function of the value of SNR^2 in that detector (see Sec. III J of TV19).

In the O1 analysis, when estimating $\mathcal{L}(t|\mathcal{N})$, we assumed that the background rate per template is flat over all subbanks in a given bank [17]. We revisited the validity of this assumption while developing the analysis in this paper. Figure 3 shows the rank functions for the five subbanks of the bank BBH 4 as

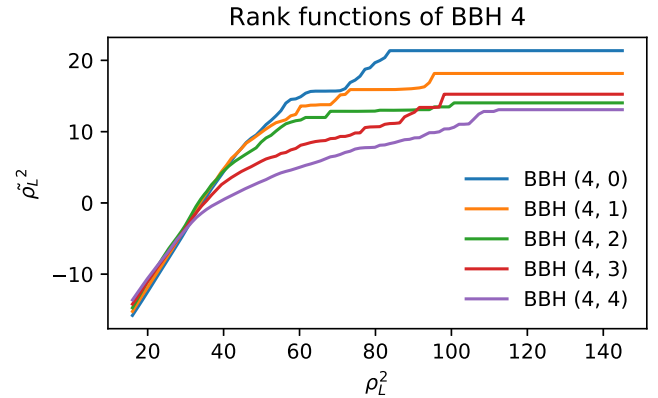


FIG. 3. Rank functions for all subbanks in bank BBH 4, as a function of the SNR^2 in the L1 detector. The quantity on the y axis is twice the logarithm of the number of triggers above a particular value of ρ_L^2 , with an offset. The different subbanks have substantially different background rates: glitches prefer the higher-mass subbanks and, even in the Gaussian part, the slopes may differ due to the different number of degrees of freedom.

a function of the SNR^2 in L1, ρ_L^2 . (These curves are completely empirically determined and not fits; the quantity on the y axis is twice the logarithm of the number of triggers above a particular value of ρ_L^2 , with an offset. See discussion in Sec. III J of TV19, and below.) The figure clearly shows that the subbanks have substantially different background rates per template. The distributions of ρ_L^2 transition from exponential (chi-squared-like) to power laws when glitches become important, and hence the rank functions flatten: this happens at progressively lower values of ρ_L^2 for higher-mass subbanks. If we do not account for this, a subbank with more glitches can disproportionately influence the background in the search.

In this search, when computing the false alarm rates of a trigger, we use the rank function of the subbank it is drawn from. This enables us to give a more local estimate of the false-alarm rate, which better accounts for the fact that templates in some regions of the search space are more prone to triggering on glitches (this effect has been noted before [22,23]). A technical point is that we compute FARs over all templates in a bank (which is the union of several subbanks), and hence we need to properly fix the normalization of the rank functions when comparing triggers from different subbanks. Appendix C describes the procedure; the rank functions shown in Fig. 3 were normalized in this way. Curiously, we see that the rank functions have different slopes in different subbanks even at low values of ρ_L^2 (in the Gaussian-noise-dominated regime); this is due to the maximization over templates when we collect triggers (see Appendix B for a derivation in a different context).

- (6) *Computing p_{astro}* : We define $\mathcal{R}(\text{event}|\mathcal{N})$ and $\mathcal{R}(\text{event}|\mathcal{S})$ to be the rates of a given event under the noise ($\mathcal{N}$) and signal ($\mathcal{S}$) hypotheses. The probability that an event is astrophysical is

$$p_{\text{astro}}(\text{event}) = \frac{\mathcal{R}(\text{event}|\mathcal{S})}{\mathcal{R}(\text{event}|\mathcal{N}) + \mathcal{R}(\text{event}|\mathcal{S})}. \quad (1)$$

We define a rate $\mathcal{R}$ for each bank to be the overall number of astrophysical events satisfying $\rho_H^2, \rho_L^2 > 16$ and $\rho_H^2 + \rho_L^2 > \rho_{\text{th}}^2$, where

$$\begin{aligned} \rho_{\text{thBBH0}}^2 &= 67, \\ \rho_{\text{thBBH1}}^2 &= 65.5, \\ \rho_{\text{thBBH2}}^2 &= 63.5, \\ \rho_{\text{thBBH3}}^2 &= 59.5, \\ \rho_{\text{thBBH4}}^2 &= 56.5. \end{aligned}$$

The rate $\mathcal{R}$ is for a hypothetical network consisting of two identical detectors, each having a sensitivity equal to the median Hanford sensitivity in the O2 run, which observe in coincidence for 118 days. $\mathcal{R}$ is assumed to be uniform across templates within the bank, regardless of which subbank they might fall in. For a given event, we have

$$\mathcal{R}(\text{event}|\mathcal{S}) = W(\text{event})\mathcal{R}, \quad (2)$$

where the factor W depends on the instantaneous sensitivities of the detectors, as well as the extrinsic parameters of the event. Note that W does not depend on the (unknown) astrophysical rate. We determine the terms in Eqs. (1) and (2) in a similar manner to that of our O1 analysis [14], but with two changes. First, we estimate the rate of producing triggers under the noise hypothesis, $\mathcal{R}(\text{event}|\mathcal{N})$, using only the background triggers in the respective subbanks that the candidates belong to. Second, we determine the rate of astrophysical events, $\mathcal{R}$, from the data itself rather than assuming it from the loudest events detected in the search.

Given a particular value of the rate $\mathcal{R}$, the likelihood of the data is

$$\mathcal{L}(\text{data}|\mathcal{R}) \propto e^{-\mathcal{R}} \times \prod_{\text{triggers}} [\mathcal{R}(\text{event}|\mathcal{N}) + W(\text{event})\mathcal{R}], \quad (3)$$

where the product is over all the triggers in the bank, including those detected originally by the LVC. Using this likelihood, we compute a posterior on $\mathcal{R}$, assuming a uniform prior $P(\mathcal{R})$ between 0 and 50. As earlier, this number is the overall number of triggers of astrophysical origin in a region in the $\rho_H^2 - \rho_L^2$ plane for a hypothetical two-detector network with each detector having the median H1 sensitivity over the O2 run, i.e., a period of 118 days. We expect this number to be of the order of the number of detections and determine it from the data itself; hence the most important requirement is that the prior be noninformative over the relevant range. For any value of the overall rate $\mathcal{R}$, we can calculate the probability that an event is of astrophysical origin; our final quoted values were obtained by marginalizing over $\mathcal{R}$:

$$p_{\text{astro}}(\text{event}) = \int_0^\infty p_{\text{astro}}(\text{event}|\mathcal{R})P(\mathcal{R})d\mathcal{R}. \quad (4)$$

TABLE I. Events already reported by the LIGO-Virgo Collaboration [1] as detected with our pipeline. The rate distributions used to compute p_{astro} are shown in Fig. 4. The maximum likelihood rates are $\mathcal{R}_{\text{max}} = 8/\text{O2}$ and $5/\text{O2}$ in banks BBH 3 and BBH 4, respectively.

Name	Bank	GPS time ^a	ρ_{H}^2	ρ_{L}^2	IFAR (O2) ^b	$\frac{W(\text{event})}{\mathcal{R}(\text{event} N)} (\text{O2})$	p_{astro}
GW170104	BBH (3, 0)	1167559936.582	85.1	104.3	$> 2 \times 10^4$	> 100	> 0.99
GW170809	BBH (3, 0)	1186302519.740	40.5	113	$> 2 \times 10^4$	> 100	> 0.99
GW170814	BBH (3, 0)	1186741861.519	90.2	170	$> 2 \times 10^4$	> 100	> 0.99
GW170818	BBH (3, 0)	1187058327.075	19.4	95.1	1.7 ^c	...	... ^c
GW170729	BBH (3, 1)	1185389807.311	62.1	53.6	$> 2 \times 10^4$	> 100	> 0.99
GW170823	BBH (3, 1)	1187529256.500	46.0	90.7	$> 2 \times 10^4$	> 100	> 0.99

^aThe times given are the “linear-free” times of the best fit templates in our bank in the Hanford detector; with this time as the origin, the phase of the template is orthogonal to shifts in time, given the fiducial PSD.

^bThe inverse false alarm rates, or IFARs, are computed within each bank and do not include any additional trial factors; our BBH analysis has five chirp-mass banks. The IFAR is given in terms of “O2” instead of physical time, since the ranking statistic includes a time-dependent volumetric correction factor to account for the significant and systematic changes in the network’s sensitivity over the run. If the network’s sensitivity were constant during the observing run, the unit “O2” ≈ 118 days.

^cSee discussion in Sec. IV.

IV. RESULTS ON THE PREVIOUSLY REPORTED EVENTS

Table I summarizes our pipeline’s results for the O2 events published by the LVC [1]. We detect all previously reported BBH events except for GW170608, for which the LVC did not provide the Hanford data in their bulk data release, and thus that time was not part of our coincidence search. We only report results from our BBH search in this paper, so we exclude the binary neutron star GW170817 from our results. Nearly all of the LVC events have only an upper limit for the FAR of $1/(20000\text{O2})$. All of these events are certainly astrophysical sources with a lower limit on $p_{\text{astro}} \geq 0.99$.

An interesting special case is GW170818, which was not found by the PyCBC pipeline, and is deemed potentially interesting but not confirmed by the GstLAL pipeline using Hanford and Livingston alone [1]. It was subsequently detected with high confidence by GstLAL when Virgo data were included; we did not analyze Virgo for the search reported in this paper. Because of the low score in H1, this is close to a single-detector trigger, and hence the detection heavily relies on our understanding of the background distribution in the vicinity of the L1 SNR². Empirically, our background distributions do not have any L1 triggers that are louder than this event. In the regime relevant to triggers like GW170818, the background is dominated by only a few loud glitches, and hence the L1 ranking score, $\tilde{\rho}_{\text{L}}^2$, that we use to define our test statistic saturates (see Fig. 3) and comes with significant error bars. Moreover, we cannot compute the *local* probability density of the background in the $\rho_{\text{H}}^2 - \rho_{\text{L}}^2$ plane without extrapolating from lower values, and this is a vital ingredient in the calculation of p_{astro} [see Eq. (1)]. Hence we need a more careful analysis to reliably assess the FAR and probabilities of astrophysical origin of such events. We defer this analysis to a follow-up paper, in which we will show that such events can be detected using the Livingston and Hanford detectors alone. In this

subsequent work, we will also report the results of a search for similar events, i.e., events that are loud in one detector (and saturate the rank score), and that either have low scores in the other detectors or have no coincident data.

The following technical details about the quantities we report in Table I (and subsequent tables) are worth noting for clarity:

- (i) Our false alarm rates (FARs) are calculated per bank and do not include any additional trials factors. Hence, the reader should interpret these numbers in the context of the set of banks used in the search. We adopt this convention since (a) FARs reported this way are invariant even if we subsequently expand the search space (i.e., add extra banks), and (b) it makes it convenient for readers to apply their own priors when computing the look-elsewhere correction, or the trials factor (i.e., the interpretation of a particular FAR is different in regions of parameter space known to have larger astrophysical rates).
- (ii) We report the linear-free times of the best fit templates in the template bank in the Hanford detector. The linear-free time is defined as follows: suppose that for a particular set of intrinsic parameters, we have a frequency domain template $h(f) = A(f) \exp[i\psi(f)]$, where the phase $\psi(f)$ is unwrapped as a function of frequency. Suppose through the inverse Fourier transform, we have a time-domain template $h(t) = \int df h(f) \exp(i2\pi ft)$, defined on the domain $[0, T]$. The linear-free time of the template $h(t)$ occurs at $t = T - t_0$, where t_0 satisfies the equations

$$\begin{bmatrix} \langle f, f \rangle & \langle f, 1 \rangle \\ \langle f, 1 \rangle & \langle 1, 1 \rangle \end{bmatrix} \begin{bmatrix} 2\pi t_0 \\ \psi_0 \end{bmatrix} = - \begin{bmatrix} \langle \psi, f \rangle \\ \langle \psi, 1 \rangle \end{bmatrix}, \quad \text{with} \quad (5)$$

$$\langle a(f), b(f) \rangle = \int df \frac{A^2(f)}{S(f)} a(f) b(f). \quad (6)$$

TABLE II. New events with astrophysical probability $> 50\%$ in all of the BBH banks. The rate distributions used to compute p_{astro} are shown in Fig. 4, and the maximum-likelihood rates in banks BBH 3 and BBH 4 are $\mathcal{R}_{\text{max}} = 8/\text{O2}$ and $5/\text{O2}$, respectively.

Name	Bank	$\mathcal{M}^{\text{det}}(\text{M}_{\odot})$	χ_{eff}	z	GPS time ^a	ρ_{H}^2	ρ_{L}^2	IFAR (O2) ^b	$\frac{W(\text{event})}{\mathcal{R}(\text{event} N)}(\text{O2})$	p_{astro}
GW170121	BBH (3, 0)	29^{+4}_{-3}	$-0.3^{+0.3}_{-0.3}$	$0.24^{+0.14}_{-0.13}$	1169069154.565	29.4	89.7	2.8×10^3	> 30	> 0.99
GW170304	BBH (4, 0)	47^{+8}_{-7}	$0.2^{+0.3}_{-0.3}$	$0.5^{+0.2}_{-0.2}$	1172680691.356	24.9	55.9	377	13.6	0.985
GW170727	BBH (4, 0)	42^{+6}_{-6}	$-0.1^{+0.3}_{-0.3}$	$0.43^{+0.18}_{-0.17}$	1185152688.019	25.4	53.5	370	11.8	0.98
GW170425	BBH (4, 0)	47^{+26}_{-10}	$0.0^{+0.4}_{-0.5}$	$0.5^{+0.4}_{-0.3}$	1177134832.178	28.6	37.5	15	0.65	0.77
GW170202	BBH (3, 0)	$21.6^{+4.2}_{-1.4}$	$-0.2^{+0.4}_{-0.3}$	$0.27^{+0.13}_{-0.12}$	1170079035.715	26.5	41.7	6.3	0.25	0.68
GW170403	BBH (4, 1)	48^{+9}_{-7}	$-0.7^{+0.5}_{-0.3}$	$0.45^{+0.22}_{-0.19}$	1175295989.221	31.3	31.0	4.7	0.23	0.56

^aThe times given are the linear-free times of the best fit templates in our bank in the Hanford detector; with this time as the origin, the phase of the template is orthogonal to shifts in time, given the fiducial PSD.

^bThe inverse false alarm rates, or IFARs, are computed within each bank, and do not include any additional trials factors; our BBH analysis has five chirp-mass banks. The IFAR is given in terms of “O2” instead of physical time, since the ranking statistic includes a time-dependent volumetric correction factor to account for the significant and systematic changes in the network’s sensitivity over the run. If the network’s sensitivity were constant during the observing run, the unit “O2” ≈ 118 days.

In the above equation, $S(f)$ is the PSD of the detector noise. Intuitively, after applying the parameter-dependent shift t_0 to the template, the measurements of arrival time and intrinsic parameters are uncorrelated in a single detector (see related discussion in Ref. [20]). When using Eqs. (5) and (6), i.e., only to define t_0 , we used the same fiducial PSD that we used to construct the template bank as described in Sec. III. For the short waveforms in the heavier binary black hole banks, the linear-free time is practically close to the merger time.

V. OVERVIEW OF THE NEWLY DISCOVERED BBH MERGERS

Table II summarizes the basic properties of the newly discovered events: their parameters, inverse FAR, or IFAR, and estimated probabilities of being of astrophysical origin, p_{astro} . We need the rate of astrophysical events in the detector, $\mathcal{R}$, to calculate these values of p_{astro} , and hence, we also report the values of the factor W [from Eq. (2)], so that external sources of information about the rate (e.g., future runs) can easily be incorporated if needed.

We derive a distribution for $\mathcal{R}$ from all the events we detected (including those already detected by the LVC analyses), and at the same time estimate the p_{astro} of each event, using the procedure described in Sec. III. Figure 4 shows the posteriors on the rates for chirp-mass banks BBH 3 and BBH 4; the values of $p_{\text{astro}}(\text{event})$ quoted in Tables II and II were marginalized over these posteriors. As is apparent from Table II, all of our detections are in the high-mass region of parameter space, which is covered by the banks BBH 3 and BBH 4. Hence, curves analogous to those in Fig. 4 for the other banks only yield upper bounds on the respective rates (of the mergers detected in previous runs, GW151012 and GW151226 lie in banks BBH 2 and BBH 1, respectively).

As in Table I, the IFARs in Table II are calculated per bank and do not include any additional trials factors. The criterion in Table II is based on the value of p_{astro} , but all the events have significantly higher values of the “per-bank” IFAR than the threshold of 1 in 30 days, or ~ 0.25 O2, adopted in GWTC-1 [1] (i.e., their IFARs exceed 0.25 O2 even if divided by a trials factor of 5).

Figure 5 shows the spectrograms for segments of data of length 1 s around the GPS times of the events in the Hanford and Livingston detectors. These spectrograms show no obvious evidence of glitches in the immediate vicinity of the events, and are visually consistent with how binary black hole merger signals with the reported values of

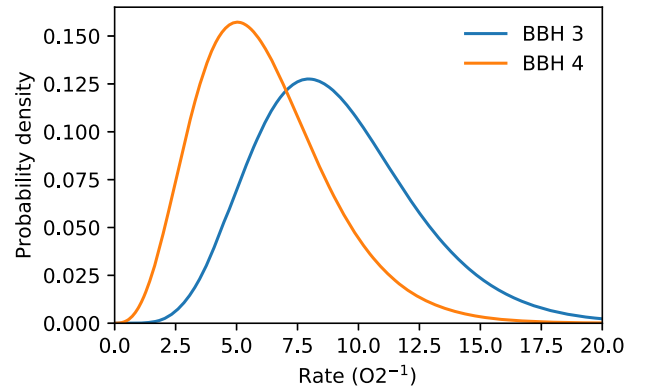


FIG. 4. Rates inferred from all the events. The rates are simplistic and bank-specific observed rates for black hole mergers, which we only use to determine the p_{astro} of the events. The rates are defined as the astrophysical occurrence rate of signals satisfying $\rho_{\text{H}}^2, \rho_{\text{L}}^2 > 16$ and $\rho_{\text{H}}^2 + \rho_{\text{L}}^2 > \rho_{\text{th}}^2$, where $\rho_{\text{th}}^2 = 60$ for BBH 3 and $\rho_{\text{th}}^2 = 57$ for BBH 4. The assumed network consists of two identical detectors with median Hanford sensitivity, observing in coincidence for 118 days. More careful analysis is required in order to infer astrophysically meaningful volumetric rates.

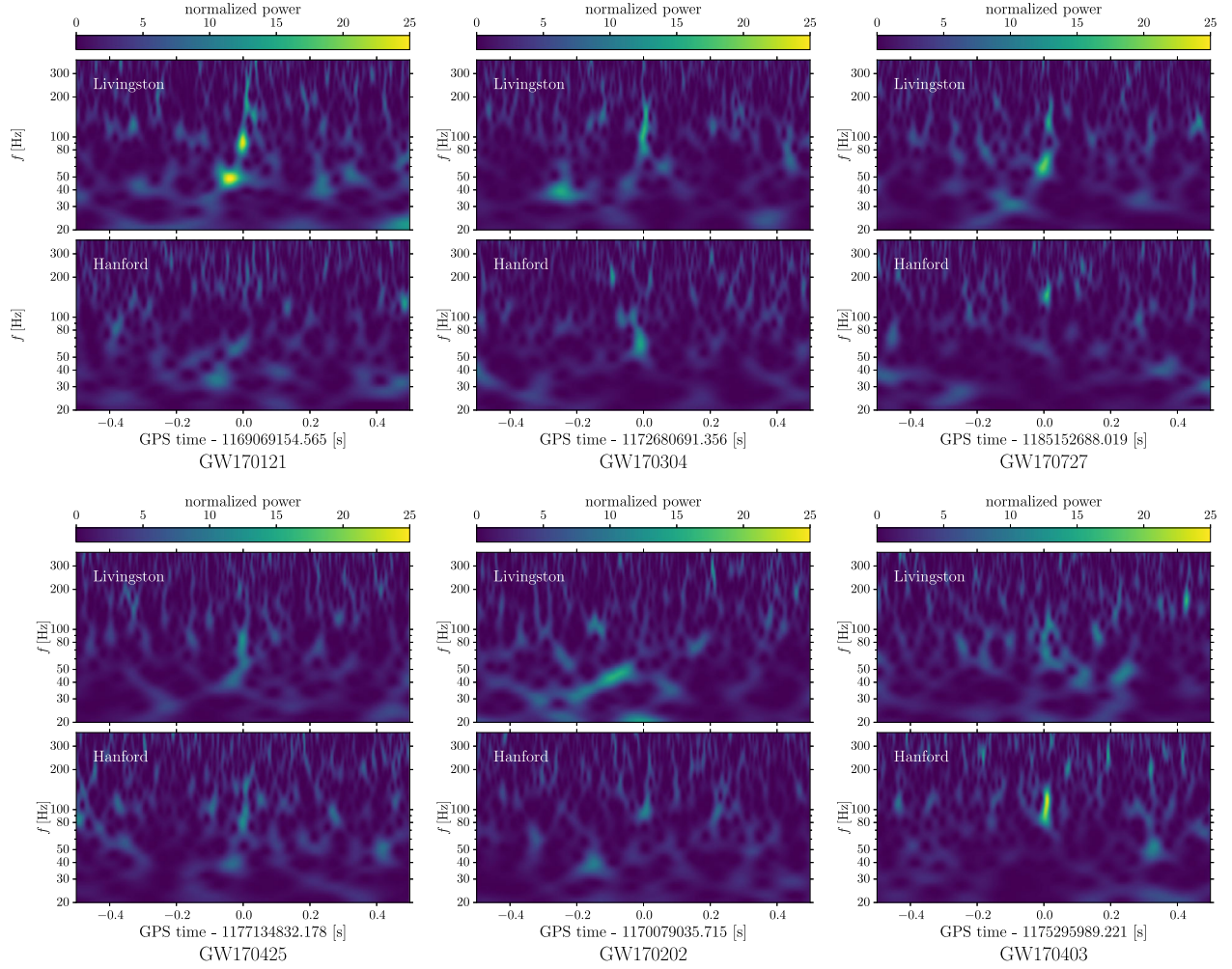


FIG. 5. The various panels show spectrograms for 1 s segments of data around the GPS times of the events in the Livingston and Hanford detectors. In each figure, the data shown was whitened with a filter constructed using the PSD measured from a 4096 s segment of data containing the event but not corrected for any short-term PSD variations within.

SNR would look in the presence of additive Gaussian noise. Note that visual inspection of spectrograms is not a particularly good or even a well-defined way to identify either glitches or signals in the data: our pipeline automatically flags bad segments of data using mathematically well-defined test statistics (see TV19 for details), and looks for signals using matched filtering.

Appendix A includes the posteriors for the parameters of all the new events. Figure 1 places these events in the context of the previous LVC events, as well as the one we reported in Ref. [18], by showing their distribution in the plane of source-frame total mass and the effective spin parameter χ_{eff} . In the remainder of this section, we briefly comment on the properties of each of the newly found events.

- (i) GW170121: This event has the lowest FAR among those not reported by the LVC ($\text{IFAR} \approx 2.8 \times 10^3$ O2), and $p_{\text{astro}} > 0.99$. The posterior distribution of χ_{eff} has most of its support at negative values, $\chi_{\text{eff}} > 0$ being ruled out at the 95% confidence level. The chirp

mass, mass ratio, and redshift of this event are similar to those of the heavy BBHs reported by the LVC. Its redshift is $z \sim 0.24$.

- (ii) GW170304 and GW170727: These two events have $\text{IFAR} \sim 370$ O2 and $p_{\text{astro}} \approx 0.98$. Their masses and spins are similar to those of the heavy BBHs detected by the LVC. Both events are consistent with zero χ_{eff} and are on the massive end of the population. They have relatively high redshifts $z \sim 0.5$ and 0.43 , respectively.
- (iii) GW170425: This candidate has $p_{\text{astro}} \approx 0.77$ and $\text{IFAR} \approx 29$ O2. Its inferred parameters are similar to those of the heavy BBHs reported by the LVC; the effective spin χ_{eff} is consistent with zero. The posterior distribution has a tail extending to large values for the masses. Its inferred redshift is large, $z \sim 0.5$.
- (iv) GW170202: This candidate has $p_{\text{astro}} \approx 0.7$ and $\text{IFAR} \approx 6$ O2. The masses and the spins are similar

TABLE III. Subthreshold candidates with astrophysical probability above 10% in all of the BBH banks. The rate distributions used to compute p_{astro} are shown in Fig. 4, and the maximum-likelihood rates in banks BBH 3 and BBH 4 are $\mathcal{R}_{\text{max}} = 8/\text{O2}$ and $5/\text{O2}$, respectively.

Bank	GPS time ^a	ρ_{H}^2	ρ_{L}^2	IFAR (O2) ^b	$\frac{W(\text{event})}{\mathcal{R}(\text{event} \mathcal{N})}$ (O2)	p_{astro}
BBH (4, 1)	1172487817.477	48.6	19.1	0.82	0.147	0.45
BBH (3, 0)	1170914187.455	20.4	41.4	0.43	0.044	0.28
BBH (3, 1)	1172449151.468	29.5	32.4	0.31	0.025	0.18
BBH (4, 0)	1174138338.385	37.1	28.4	0.62	0.034	0.17
BBH (3, 0)	1171863216.108	46.5	21.6	0.27	0.016	0.125
BBH (3, 1)	1187176593.222	20.3	42.0	0.2	0.014	0.12
BBH (3, 0)	1182674889.044	34.1	28.7	0.23	0.016	0.12
BBH (3, 1)	1171410777.200	40.8	21.0	0.18	0.014	0.11

^aThe times given are the linear-free times of the best fit templates in our bank in the Hanford detector; with this time as the origin, the phase of the template is orthogonal to shifts in time, given the fiducial PSD.

^bThe inverse false alarm rates, or IFARs, are computed within each bank and do not include any additional trials factors; our BBH analysis has five chirp-mass banks. The IFAR is given in terms of “O2” instead of physical time, since the ranking statistic includes a time-dependent volumetric correction factor to account for the significant and systematic changes in the network’s sensitivity over the run. If the network’s sensitivity were constant during the observing run, the unit “O2” ≈ 118 days.

to those of the heavy LVC BBHs. It is found in the bank with the largest number of secure detections (BBH 3). It has a bimodal posterior, in which the solution with lower masses has a more negative spin, and is located closer. The inferred redshift is $z \sim 0.27$.

- (v) GW170403: This candidate has $\text{IFAR} \approx 5 \text{ O2}$ and $p_{\text{astro}} \approx 0.55$; this is close to the threshold $p_{\text{astro}} = 0.5$ to make it into a list of detections (as defined in Ref. [1]). The inferred redshift is $z \sim 0.45$. Interestingly, the posterior for χ_{eff} is inconsistent with positive values.

In addition to these events, we list in Table III the subthreshold triggers of our search, defined as those with $0.1 < p_{\text{astro}} < 0.5$. The sum of the p_{astro} of the events in this list exceeds unity; in fact, a candidate in bank BBH (4, 1) has $p_{\text{astro}} \approx 0.45$, which is close to the detection threshold (though it has a relatively high $\text{IFAR} \approx 0.8 \text{ O2}$). It is possible that an improved analysis, or rate estimate, can push some of these candidates above the detection threshold. Note that we have a threshold of $p_{\text{astro}} > 0.1$ for candidates to appear in Table III. The set of loudest triggers in banks BBH 2, BBH 1, and BBH 0 did not pass this cut.

VI. SENSITIVITY OF OUR PIPELINE

In the previous section, we described several additional events we detected that are not in the catalog of events published by the LVC. All of these events pass the thresholds for detection in GWTC-1 [1] (their FARs are above the threshold of 1 in 30 days, even accounting for the five banks in our BBH search, or 11 banks in a hypothetical binary neutron star and neutron-star–black-hole search [20], in the extreme scenario in which searches for different kinds of astrophysical systems are considered together).

The values of the FAR and p_{astro} of the events come with error bars, and for the same event, their values can fluctuate even between two searches that use the same algorithms due to choices within. Given that searches operate with thresholds, the presence or absence of a near-threshold event in one search, *by itself*, is not proof of a difference in sensitivity. However, equally sensitive searches should agree on the overall population of events, and, on average, assign comparable values of FAR and p_{astro} to comparable events. Our results in Sec. IV show that we assign systematically lower values of FAR to the events discovered by the LVC pipelines in general, and in particular, for GW170729, the least secure event in GWTC-1. Moreover, we have a substantial number of events that individually clear the thresholds for detection; while it is certainly possible that some of the events with lower values of p_{astro} could have benefited from upward fluctuations in their test statistics, it is hard to explain the entire population away this way. These two effects suggest that our analysis has a substantially larger sensitive volume within the search space that we defined in Sec. III.

In an idealized case, a simple way to compare sensitive volumes would be to inject a large number of signals in the data and measure the fraction recovered above the appropriate thresholds in the test statistics adopted. We will present the results of such a systematic injection campaign in a future paper; pending this, we would like to obtain a simple estimate of the additional sensitivity. In the rest of this section, we will consider the population of loud astrophysical events in this data akin to a common set of injected signals and get a simple “back-of-the-envelope” level estimate of the order of magnitude of the sensitivity change. Since there are relatively few events, this is necessarily highly uncertain, and hence the results should not be overinterpreted.

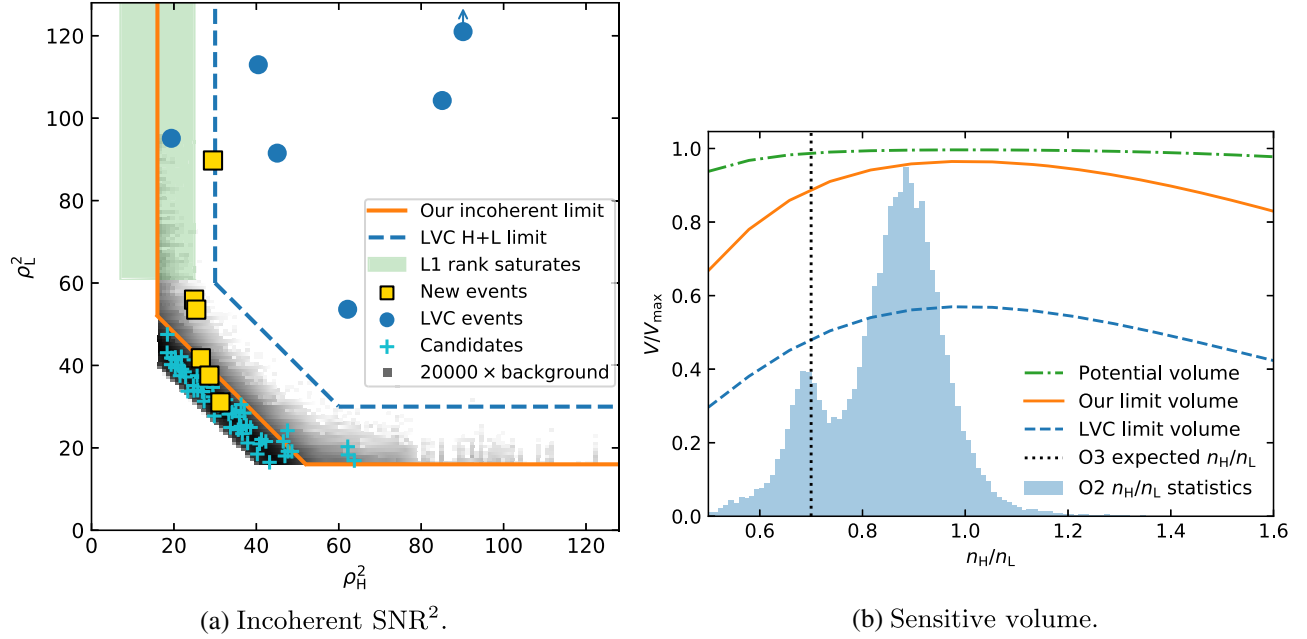


FIG. 6. (a) Incoherent Hanford and Livingston SNR^2 for coincident and background triggers (computed using 2×10^4 time slides), for all the subbanks where there are events. The blue and orange lines are approximate incoherent detection limits for the LVC and the current analysis, respectively, restricted to using Hanford and Livingston data only (see text for caveats). GW170814 has $\rho_L^2 = 170$, higher than shown here (indicated with an arrow). GW170608 is not shown; see Sec. IV. (b) The lines show the (incoherent) volume probed by different analyses as a function of the ratio of the Hanford and Livingston detector sensitivities, at fixed network total sensitivity. The orange (solid) and blue (dashed) curves show the estimated volume probed by the present and LVC analyses, respectively. The green, dash-dotted curve shows the potential volume that can be opened up by analyzing interesting single-detector triggers. The shaded histogram shows the distribution of $V \times T$, i.e., the product of the sensitive volume and the time, as a function of the sensitivity ratios between the Hanford and Livingston detectors in the O2 run, as measured by our analysis. The O3 run, as reported in Ref. [24], has begun with a sensitivity ratio $n_H/n_L \approx 0.7$.

Figure 6(a) shows the background triggers we collected using 20 000 time slides in those BBH subbanks in which all the events considered in this work, from both the LVC and our analysis, reside. This figure does not include the BBHs from the O1 run (GW150914, GW151012, GW151216, GW151226), nor GW170608, which was not included in the bulk data release we analyzed. This figure is not intended as a demonstration of how we compute the FAR or p_{astro} for particular events: first, it shows ρ_H^2 and ρ_L^2 , i.e., the incoherent H1 and L1 SNR^2 , while we compute the FAR using a coherent score that takes into account the time delays and the relative phases of the triggers, and the differing detector sensitivities; second, we estimate the FAR and the p_{astro} for a particular event using the background in its chirp-mass bank and subbank, respectively. We include this figure only to easily visualize the sensitive volume.

It is clear from Fig. 6 that our pipeline has a substantially lower background in the relevant region: for example, we see no background triggers within the sensitive region of the LVC pipelines. All but one of the LVC reported events have values of ρ_L^2 that are so large that we do not have even single-detector background triggers at their level (this is a consequence of our data-cleaning procedure, as well as our

signal-quality vetoes). The only exception to this is GW170729, with $(\rho_H^2, \rho_L^2) = (62, 53)$. This event had FARs of 0.2 yr^{-1} and 1.36 yr^{-1} in the *GstLAL* and *PyCBC* pipelines, respectively, but we have no background in its vicinity even incoherently (i.e., allowing for arbitrary phases, time delays, and sensitivity ratios of the two detectors).

To calculate sensitive volumes, we need to estimate the detection limits of the analysis. With the above caveat on the validity of thresholds based on incoherent SNR^2 , the solid and dashed lines show the approximate detection thresholds that we judged as appropriate for different analyses. The detection thresholds shown for the LVC catalog are approximate and conservative; they err on the side of reporting a better sensitivity. At the single-detector level, we set the threshold to $\rho_H^2, \rho_L^2 > 30$ by the non-detection of GW170121. We set the minimum network $\text{SNR}^2 = \rho_H^2 + \rho_L^2 > 90$ by scaling the reported FAR of GW170729 to $1/\text{O2}$, and rounding down. We approximated our incoherent limit as $\rho_H^2 + \rho_L^2 > 68$ and $\rho_H^2, \rho_L^2 > 16$, based on our cuts and the FAR we would assign to events at this level given our search background.

The difference in the detection limits, at the same threshold on FAR, maps to a difference in the sensitive

volume between the searches. The size of this difference depends on the ratio of the sensitivities of the two detectors we analyze, i.e., H1 and L1. The sensitivities n_H and n_L are proportional to the SNR with which a gravitational wave signal with a given strain amplitude is measured by H1 and L1, respectively. Figure 6(b) shows the volume as a function of the sensitivity ratio (scaled such that a volume of $V_{\max}$ corresponds to detecting all events with $\rho_H^2 + \rho_L^2 > 68$); the solid orange and the dashed blue lines are for the cuts shown in Fig. 6(a). The curves were obtained by randomizing the angular locations and inclinations of a large number of mergers on the sky and recording the incoherent scores (including the stochastic noise contribution) at the detectors with a given sensitivity ratio. We see that the gain between the blue and orange lines depends significantly on the relative sensitivities of the H1 and L1 detectors. The filled histogram in Fig. 6(b) shows the distribution of the ratio of the Hanford and Livingston sensitivities in our analysis of the O2 run, weighted by the momentary spacetime volume. We see that the orange curve gains approximately a factor of 2 in sensitive volume over the relevant range of sensitivities. As emphasized in the beginning of this section, this is only a rough, order-of-magnitude estimate of the relative sensitivity using a small number of events. This number comes with significant uncertainty and, as such, should not be overinterpreted.

Also of particular interest is the limit in which the SNR is much larger in one detector than in the other. For part of the O2 run, the Livingston detector was substantially more sensitive than the Hanford one, and hence there is a substantial phase-space volume for astrophysical signals to have disparate SNR in the detectors. Figure 6(a) shows that there is substantially less background in L1 in the high SNR regime (see the teal shaded region). The few background events in this region come from the same small number of loud events in L1 matching with Gaussian fluctuations in H1 at different time shifts, and hence the ranking function $\tilde{\rho}_L^2$ saturates and is severely affected by Poisson noise (see Fig. 3). We need a different analysis to estimate a meaningful FAR in this regime; the green, dash-dotted curve in Fig. 6(b) shows that we can gain a nontrivial amount of sensitive volume if we open up this region. We will study this regime in more detail in a subsequent paper, in which we will introduce a formalism for searching for events and assessing the FAR in this region. Notably, the LVC event GW170818, which was detected using Virgo data, belongs to this category.

VII. CONCLUSIONS

In this paper we presented the results of our search for BBHs in the data from the O2 observing run of advanced LIGO, using the methods introduced in our work in Ref. [14]. We report six new events above the detection

thresholds defined by the LVC (in terms of FAR and p_{astro}), three of whom have probability $p_{\text{astro}} > 0.98$ of being of astrophysical origin. Interestingly, all the new events are in banks BBH 3 and BBH 4 (our heavy chirp-mass banks), as are most of the ones reported by the LVC.

The most significant new event (GW170121) prefers negative χ_{eff} and is inconsistent with positive values at the 95% level. The most marginal candidate event (GW170403), with $p_{\text{astro}} \sim 0.5$, is inconsistent with zero or positive χ_{eff} . The spin of the merging BBHs is an important discriminator between formation channels [25]. Hence, the new events presented in this work can throw light on the mechanisms by which BBHs are assembled.

More generally, with the increased number of events, the clear next step is to perform a population analysis that accounts for selection biases, which will map out the distribution of the intrinsic parameters of the mergers. In particular, including new events in population analyses can significantly inform us about the dependence of the merger rate on mass and redshift.

The LVC recently started their third observing run (O3), and several new detections are expected in the near future. The new events we report in this paper show that there will be additional information in the LIGO and Virgo data in addition to what the pipelines used by the LVC currently extract. The development of our pipeline has been facilitated by access to the public O1 and O2 data, as well as the LIGO Algorithm Library [26]. We thank the LVC for releasing the data and tools to the community. We hope that data from current and future runs can be made available quickly to incentivize external groups to develop new analysis methods and maximize the scientific yield of the LVC data.

ACKNOWLEDGMENTS

This research made use of data, software, and/or web tools obtained from the Gravitational Wave Open Science Center (<https://www.gw-openscience.org>), a service of LIGO Laboratory, the LIGO Scientific Collaboration, and the Virgo Collaboration. LIGO is funded by the U.S. National Science Foundation. Virgo is funded by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale della Fisica Nucleare (INFN), and the Dutch Nikhef, with contributions by Polish and Hungarian institutes. T. V. acknowledges support by the Friends of the Institute for Advanced Study. B. Z. acknowledges the support of The Peter Svennilson Membership fund. L. D. acknowledges the support from the Raymond and Beverly Sackler Foundation Fund. M. Z. is supported by NSF Grants No. AST-1409709, No. PHY-1521097, and No. PHY-1820775, the Canadian Institute for Advanced Research (CIFAR) program on Gravity and the Extreme Universe, and the Simons Foundation Modern Inflationary Cosmology initiative.

APPENDIX A: PROPERTIES OF THE NEW EVENTS

We performed a full parameter estimation for all the new events presented in this paper. We coherently analyzed the data from all available detectors (including Virgo), in

contrast to the search, which only used Hanford and Livingston data. We use a likelihood model that assumes the data are the sum of a gravitational wave signal and additive Gaussian random noise and a prior that is uniform in the intrinsic source-frame parameters m_1 and m_2 , the

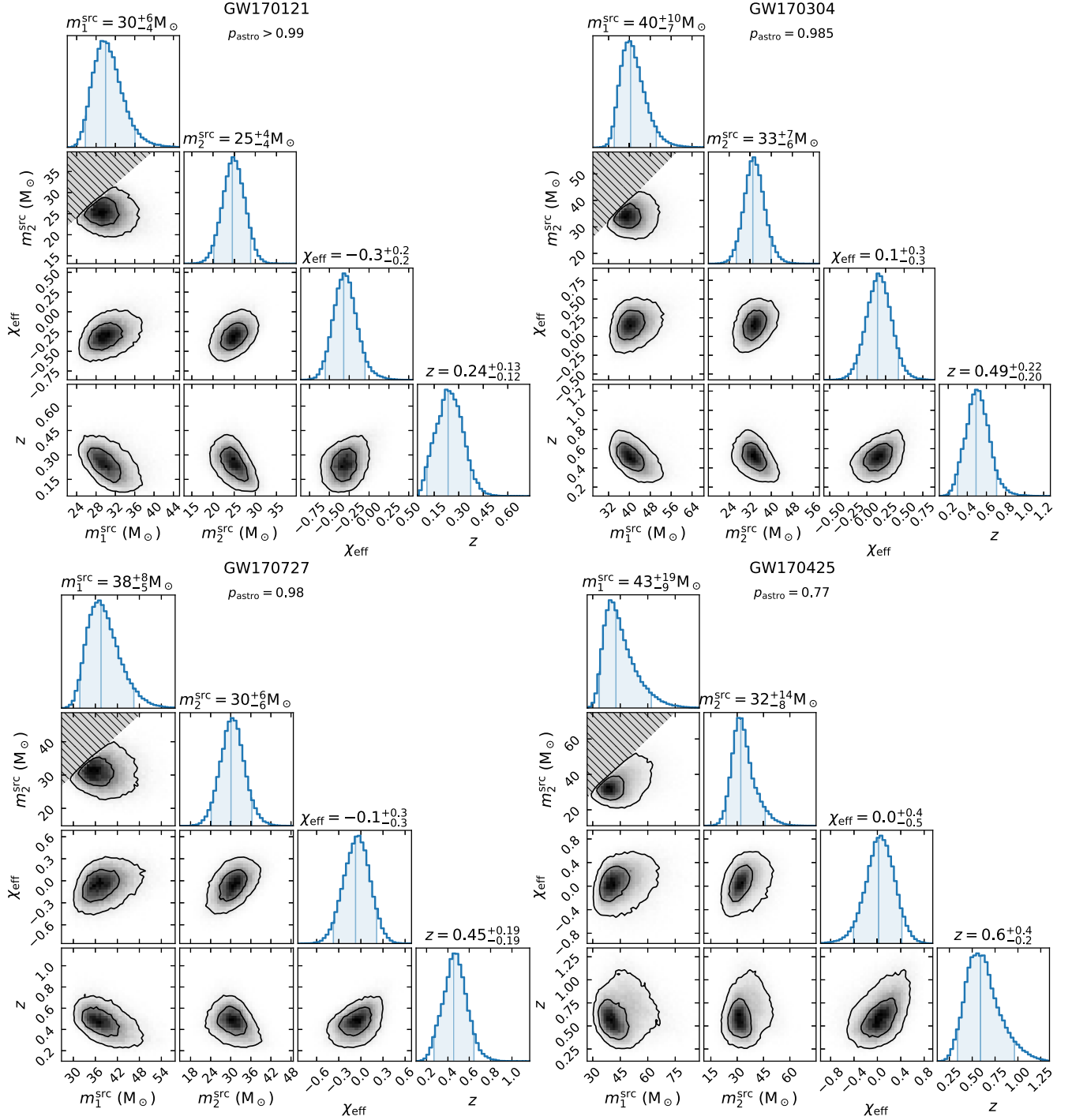


FIG. 7. Marginalized posteriors for the new events reported in this work (continued in Fig. 8). Two-dimensional contours enclose 50% and 90% of the distribution. In the one-dimensional posteriors, vertical lines show the 0.05, 0.5, and 0.95 quantiles. The samples were generated using the IMRPhenomD approximant [27] and a prior that is uniform in detector-frame m_1 , m_2 , χ_{eff} , and luminosity volume [18].

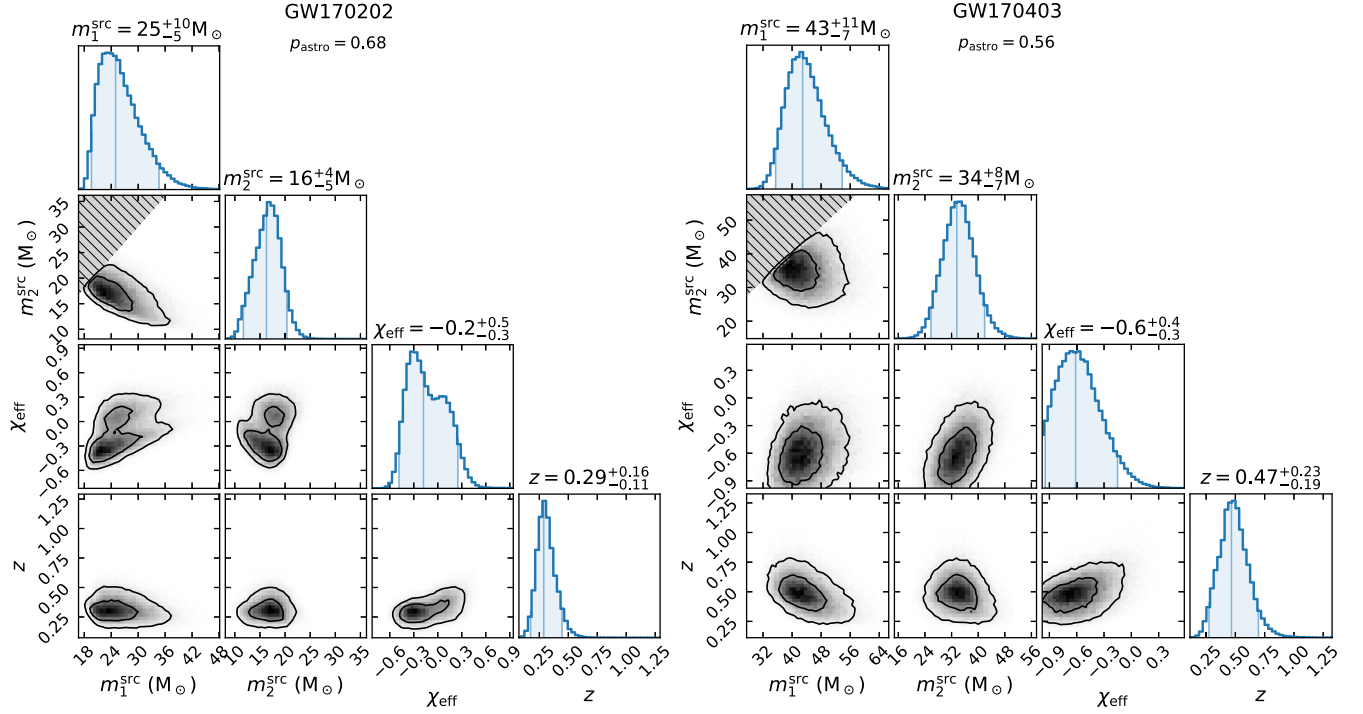


FIG. 8. Continuation of Fig. 7 displaying marginalized posteriors for GW170202 and GW170403.

effective spin parameter χ_{eff} , and the luminosity volume (as detailed in Ref. [18]). We use the IMRPhenomD approximant [27] to generate waveforms, the relative binning method [28] to evaluate the likelihood, and PyMultiNest [29] to generate samples from the posteriors.

Figures 7 and 8 show the posterior distributions of detector-frame chirp mass, mass ratio, effective spin, and redshift for the new events reported with $p_{\text{astro}} > 0.5$, marginalized over the extrinsic parameters

of the mergers such as their sky location and inclination. Figure 9 shows the posteriors for the location of all the events on the sky, marginalized in turn over intrinsic parameters and inclination: we only show the 50% and 90% contours in order to present all the events on the same plot. Samples from the full posterior distribution (i.e., the joint distribution of the intrinsic and extrinsic parameters) are available at https://github.com/jroulet/O2_samples.

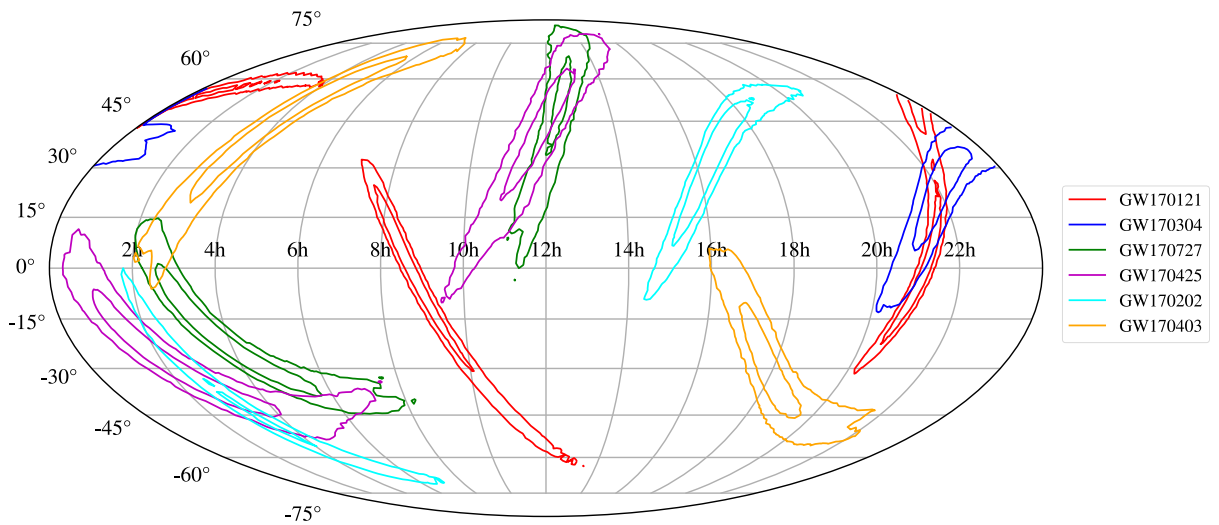


FIG. 9. Contours of 90% and 50% credible regions for the sky locations of the six new BBH events shown in the Mollweide projection.

APPENDIX B: POPULATING EVENTS IN SUBBANKS AND BANKS

In our analysis, we ultimately assign events to a single chirp-mass bank. We compute false alarm rates for an event within a particular bank by comparing its corresponding trigger to the triggers from background events (i.e., those generated by noise) in that bank. However, a single signal or noise transient produces triggers in several subbanks, and even across chirp-mass banks. If we do not properly assign events (both real and background) to banks, we run the risk of losing sensitivity in our search, and hence we should carefully choose our criteria for bank assignment.¹

We would like to perform the analysis as blindly as possible, and hence would like to avoid tweaking criteria after touching the data. Since we are analyzing a new dataset in this work, this gives us an opportunity to update the bank assignment procedure from the one we used for the O1 data (step 6 of Sec. II of TV19). We begin by considering the distribution of the triggers that a particular signal in the data produces across the banks. This distribution depends on the number of templates placed in each bank (effectively their density in parameter space), as well as the effective dimensionality of the parameter space in each bank/subbank (the dimensionality has a concrete meaning in the c_α parameter space; see Ref. [20]). Given this understanding, we use the set of triggers that an event throws up in a given bank to compute a net likelihood, and use this as a discriminator for assigning events to banks. In the rest of this Appendix, we describe the details of this procedure.

For simplicity, let us start with triggers in a single detector. Let d denote the strain data, A be the signal amplitude, and Θ be the other parameters of the signal: these include the coefficients $c_{\alpha,i}$ in subbank i of chirp-mass bank B , and the phase and the merger time. The likelihood of the subbank i under the signal hypothesis is

$$\mathcal{L}(d|i, B, S) = \sum \iint d\Theta dA p(\Theta) p(A) \mathcal{L}(d|A, \Theta), \quad (\text{B1})$$

where the sum runs over the grid points in $c_{\alpha,i}$ and the integral is over continuous parameters (time and phase). We can view the sum over the coefficients $c_{\alpha,i}$ as a Riemann sum for an integral over the underlying continuous space, $\mathcal{V}$, and thus approximate Eq. (B1) as

$$\mathcal{L}(d|i, B, S) \approx \frac{1}{(\Delta c_{\alpha,i})^{n_{c_{\alpha,i}}}} \iint d\Theta dA p(A) p(\Theta) \mathcal{L}(d|A, \Theta), \quad (\text{B2})$$

¹This is not unique to our search. For an extreme example, GW150914 produced triggers across the entire binary black hole parameter space in the LVC search [30]. If one of the fainter triggers had been picked as representing the event, it would have been assigned a different false alarm rate.

where $\Delta c_{\alpha,i}$ and $n_{c_{\alpha,i}}$ are the spacing and dimensionality of the template grid in subbank i . The factor in front of the integral is the volume per template in the discrete grid.

We adopt the following assumptions:

- (1) In the prior $p(\Theta)$, astrophysical signals are equally likely to occur in two different chirp-mass banks.
- (2) The astrophysical rate per template is uniform within a chirp-mass bank, i.e., given that a signal occurs within bank B , the probability that it occurs with a given template equals $1/N_{\text{temp},B} = 1/\sum_i N_{\text{temp},i}$, where $N_{\text{temp},B}$ and $N_{\text{temp},i}$ are the number of templates in the bank B and sub-bank i , respectively (we make the same assumption when estimating the detector rate of events in Sec. III).
- (3) The integral in Eq. (B2) receives most its contribution from around the best fit parameters, Θ_* , and amplitude A_* , which is valid in the limit of high SNR.

The merger time and phase are uniformly distributed within their ranges, while the amplitude has a prior distribution $p(A) \sim 1/A^4$ in a Euclidean universe. The integrand $\mathcal{L}(d|A, \Theta)$ has the form

$$\mathcal{L}(d|A, \Theta) = \exp \left[-\frac{\langle d - At(\Theta) | d - At(\Theta) \rangle}{2} \right], \quad (\text{B3})$$

where $t(\Theta)$ is the template and the inner product is weighted by the inverse PSD. Next, we integrate over the amplitude in Eq. (B2). Under assumption 3 above, this gives us a prefactor A_*^{-4} and simplifies the integrand to

$$\mathcal{L}(d|\Theta) \propto \exp \left[\frac{\langle d | t(\Theta) \rangle^2}{2 \langle t(\Theta) | t(\Theta) \rangle} \right], \quad (\text{B4})$$

where we have removed a term that does not depend on the templates (this step is identical to the standard derivation of the $\mathcal{F}$ statistic [31]). We express the best fit amplitude as $A_* = (\rho/\rho_0)A_0$, where ρ_0 is the SNR of a merger at a fiducial distance and orientation, which captures the instantaneous sensitivity of the detector.

By assumption 3 above, the integrand in Eq. (B4) is sharply peaked around the best-fit parameters Θ_0 ; there it equals $\exp(\rho^2/2)$, where $\rho^2 = \langle A_* t(\Theta) | A_* t(\Theta) \rangle$ is the SNR². A nice feature of our template banks is that the deviation in coefficients, δc_α , directly measures the degradation in the overlap between templates [20]:

$$\langle A_* t(c_\alpha) | A_* t(c_{\alpha,0}) \rangle \approx \rho^2 \left(1 - \frac{\delta c_\alpha^2}{2} \right). \quad (\text{B5})$$

A similar relation holds for the other continuous parameters (time and phase); i.e., the degradation in overlaps is quadratic with displacement (the principal directions are some linear combinations of time and phase), and the width is inversely related to ρ . We simplify the integrand of

Eq. (B4), substitute it into Eq. (B2), and use the prior on the templates from assumptions 1 and 2 to obtain

$$\begin{aligned} \mathcal{L}(d|i, B, S) &\propto \frac{(\rho/\rho_0)^{-4}}{N_{\text{temp},B}(\Delta c_{\alpha,i})^{n_{c_{\alpha,i}}}} \iint d\delta\Theta \exp \left[\frac{\rho^2}{2} (1 - \delta\Theta^2) \right] \\ &= \frac{(\rho/\rho_0)^{-4} \exp(\rho^2/2)}{N_{\text{temp},B}(\Delta c_{\alpha,i})^{n_{c_{\alpha,i}}}} \left(\frac{2\pi}{\rho^2} \right)^{(n_{\Theta} \equiv (n_{c_{\alpha,i}}+2))/2}. \end{aligned} \quad (\text{B6})$$

When we refine coincident triggers between two detectors, the Θ contains four extra parameters apart from the template coefficients: two times, and two phases. A complication is that astrophysical signals are not uniformly distributed in the space of the time delay and relative phase, and the distribution depends on the relative sensitivities of the two detectors, $\rho_{0,H}/\rho_{0,L}$. Thus there is a nontrivial prior in the space of parameters Θ .

We make progress by noting that the likelihood in the integrand in Eq. (B2) is independent of the relative times and phases. In this case, we are greatly helped by our assumption 3 above, which tells us that all we need is to evaluate the prior at the best-fit parameters Θ_* . We combine the prior and amplitude prefactor together into a function $p(\Delta t, \Delta\phi, \rho_H^2, \rho_L^2 | \rho_{0,H}, \rho_{0,L})$, which we evaluate by Monte Carlo sampling methods.

The integral over the likelihood in Eq. (B2) can be evaluated in a similar manner as above, with the difference that now $\rho^2 = \rho_H^2 + \rho_L^2$. Thus we finally obtain

$$\begin{aligned} \mathcal{L}(d_H, d_L|i, B, S) &\propto \frac{p(\Delta t, \Delta\phi, \rho_H^2, \rho_L^2 | \rho_{0,H}, \rho_{0,L}) e^{\rho^2/2}}{N_{\text{temp},B}(\Delta c_{\alpha,i})^{n_{c_{\alpha,i}}}} \left(\frac{2\pi}{\rho^2} \right)^{(n_{c_{\alpha,i}}+4)/2}. \end{aligned} \quad (\text{B7})$$

Note that the exponent of the last term is different due to the extra degrees of freedom in the two-detector case.

We use the ratio of the likelihoods given by Eq. (B7) as a discriminator to compare triggers in the same location in different subbanks (and possibly chirp-mass banks).

APPENDIX C: DEFINITION OF RANK FUNCTIONS IN DIFFERENT SUBBANKS

We compute FARs of events by comparing them to background triggers in their chirp-mass bank. An essential ingredient in this computation is the likelihood $\mathcal{L}(t|\mathcal{N})$ for a trigger t under the noise hypothesis. Coincident triggers are produced with random time delays and relative phases between the two detectors, and the likelihood $\mathcal{L}(t|\mathcal{N})$ depends only on the incoherent SNR^2 in the two detectors

and the template in subbank i (which we denote by the set of coefficients c_α of the basis phase functions), i.e., $\mathcal{L}(t|\mathcal{N}) = P(\rho_H^2, \rho_L^2, c_\alpha, i)$. We can write

$$P(\rho_H^2, \rho_L^2, c_\alpha, i) = P(\rho_H^2, \rho_L^2, c_\alpha|i)P(i) \quad (\text{C1})$$

$$= \frac{P(\rho_H^2|i)P(\rho_L^2|i)}{N_{\text{temp},i}} P(i), \quad (\text{C2})$$

where $P(i)$ is the probability that noise produces a coincident trigger in subbank i , regardless of the SNR^2 , and $N_{\text{temp},i}$ is the number of templates in subbank i .

Equation (C2) assumes that (a) the background is flat over templates within each subbank, and (b) the triggers in different detectors are independent of each other. Both these assumptions fail to some degree (the latter happens because we refine coincident triggers and pick the best common template). One solution would be to directly estimate the probability in Eq. (C1) from the background, but in practice the many-dimensional distribution is hard to sample finely enough, and thus real coincident triggers can receive spurious penalties to their scores. We use the same assumptions to rank the time slides as well as the coincident triggers, and hence our FARs are not biased for the strategy we adopt; the price of the above assumptions is that our ranking is no longer strictly optimal.

The probability $P(i)$ is

$$P(i) = \frac{N_{\text{trig},i}}{N_{\text{trig}}}, \quad (\text{C3})$$

where $N_{\text{trig},i}$ is the number of triggers in subbank i and N_{trig} is the number of triggers summed over all subbanks.

We approximate the probability $P(\rho^2|i)$ for each detector using the same ranking function that we adopted in our previous work [14], but compute it separately for each subbank:

$$\frac{P(\rho^2|i)}{P(\rho_0^2|i)} \approx \frac{\text{Rank}(\rho^2|i)}{\text{Rank}(\rho_0^2|i)}, \quad (\text{C4})$$

where $\text{Rank}(\rho^2|i)$ is the ranking of a given trigger in its subbank (with the lowest rank given to the loudest event) and ρ_0^2 is a normalization point that we set to $\rho_0^2 \approx 30$. We estimate $P(\rho_0^2|i)$ by taking the ratio between the number of triggers in a bin around ρ_0^2 in subbank i over the total number of triggers in the subbank, i.e.,

$$P(\rho_0^2|i) \propto -\frac{1}{N_{\text{trig},i}} \frac{d}{d\rho^2} \text{Rank}(\rho^2|i)|_{\rho=\rho_0}. \quad (\text{C5})$$

- [1] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. X* **9**, 031040 (2019).
- [2] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **116**, 061102 (2016).
- [3] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **116**, 241103 (2016).
- [4] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. X* **6**, 041015 (2016).
- [5] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **118**, 221101 (2017).
- [6] B. P. Abbott *et al.*, *Astrophys. J.* **851**, L35 (2017).
- [7] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **119**, 141101 (2017).
- [8] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **119**, 161101 (2017).
- [9] K. Chatziioannou, R. Cotesta, S. Ghonge, J. Lange *et al.*, *Phys. Rev. D* **100**, 104015 (2019).
- [10] S. A. Usman, A. H. Nitz, I. W. Harry, C. M. Biwer *et al.*, *Classical Quantum Gravity* **33**, 215004 (2016).
- [11] S. Sachdev, S. Caudill, H. Fong, R. K. L. Lo *et al.*, [arXiv:1901.08580v1](https://arxiv.org/abs/1901.08580v1).
- [12] S. Klimenko, G. Vedovato, M. Drago, F. Salemi *et al.*, *Phys. Rev. D* **93**, 042004 (2016).
- [13] A. H. Nitz, C. Capano, A. B. Nielsen, S. Reyes, R. White, D. A. Brown, and B. Krishnan, *Astrophys. J.* **872**, 195 (2019).
- [14] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, *Phys. Rev. D* **100**, 023011 (2019).
- [15] Gravitational Wave Open Science Center (GWOSC), www.gw-openscience.org/O2/.
- [16] M. Vallisneri, J. Kanner, R. Williams, A. Weinstein, and B. Stephens, *J. Phys. Conf. Ser.* **610**, 012021 (2015).
- [17] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, *Phys. Rev. D* **100**, 023011 (2019).
- [18] B. Zackay, T. Venumadhav, L. Dai, J. Roulet, and M. Zaldarriaga, *Phys. Rev. D* **100**, 023007 (2019).
- [19] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044007 (2016).
- [20] J. Roulet, L. Dai, T. Venumadhav, B. Zackay, and M. Zaldarriaga, *Phys. Rev. D* **99**, 123022 (2019).
- [21] A. H. Nitz, T. Dent, T. Dal Canton, S. Fairhurst, and D. A. Brown, *Astrophys. J.* **849**, 118 (2017).
- [22] T. Dal Canton, S. Bhagwat, S. V. Dhurandhar, and A. Lundgren, *Classical Quantum Gravity* **31**, 015016 (2014).
- [23] S. Bose, S. Dhurandhar, A. Gupta, and A. Lundgren, *Phys. Rev. D* **94**, 122004 (2016).
- [24] Gravitational Wave Open Science Center detector status, www.gw-openscience.org/detector_status/day/20190412/.
- [25] B. Farr, D. E. Holz, and W. M. Farr, *Astrophys. J.* **854**, L9 (2018).
- [26] LIGO Scientific Collaboration, LIGO Algorithm Library—LALSuite, free software (GPL) (2018).
- [27] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, *Phys. Rev. D* **93**, 044007 (2016).
- [28] B. Zackay, L. Dai, and T. Venumadhav, [arXiv:1806.08792](https://arxiv.org/abs/1806.08792).
- [29] J. Buchner, A. Georgakakis, K. Nandra, L. Hsu, C. Rangel, M. Brightman, A. Merloni, M. Salvato, J. Donley, and D. Kocevski, *Astron. Astrophys.* **564**, A125 (2014).
- [30] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy *et al.*, *Phys. Rev. D* **93**, 122003 (2016).
- [31] P. Jaranowski and A. Królak, *Living Rev. Relativity* **15**, 4 (2012).