Detecting gravitational waves with disparate detector responses:
Two new binary black hole mergers

Barak Zackay[1,*] Liang Dai,[1] Tejaswi Venumadhav,[1] Javier Roulet[2] and Matias Zaldarriaga[1]

[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*

We introduce a new technique to search for gravitational wave events from compact binary mergers that produce a clear signal only in a single gravitational wave detector, and marginal signals in other detectors. Such a situation can arise when the detectors in a network have different sensitivities, or when sources have unfavorable sky locations or orientations. We start with a short list of loud single-detector triggers from regions of parameter space that are empirically unaffected by glitches (after applying signal-quality vetoes). For each of these triggers, we compute evidence for astrophysical origin from the rest of the detector network by coherently combining the likelihoods from all detectors and marginalizing over extrinsic geometric parameters. We report the discovery of two new binary black hole (BBH) mergers in the second observing run of Advanced LIGO and Virgo (O2), in addition to the ones that were reported in [B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. X **9**, 031040 (2019) and [T. Venumadhav *et al.*, Phys. Rev. D **101**, 083030 (2020)]. We estimate that the two events have false alarm rates of one in 19 years (60 O2) and one in 11 years (36 O2). One of the events, GW170817A, has primary and secondary masses $m_1^{\rm src} = 56^{+16}_{-10}\ M_\odot$ and $m_2^{\rm src} = 40^{+10}_{-11}\ M_\odot$ in the source frame. The existence of GW170817A should be very informative about the theoretically predicted upper mass gap for stellar mass black holes. Its effective spin parameter is measured to be $\chi_{\rm eff} = 0.5 \pm 0.2$, which is consistent with the tendency of the heavier detected BBH systems to have large and positive effective spin parameters. The other event, GWC170402, will be discussed thoroughly in future work.

I. INTRODUCTION

The LIGO-Virgo Collaboration (LVC) detected ten binary black hole (BBH) coalescence events during their first and second observing runs (O1 and O2) [1]. We performed an independent analysis of the publicly released O1 and O2 data, and reported seven additional BBH events in Refs. [2–4]. Several of the events we identified were recently also found in an independent search using the PyCBC analysis pipeline, which also reported a new massive BBH [5].

Figure 1 summarizes the sensitivity reach of both search efforts, in terms of the signal-to-noise ratios (SNR) in the Hanford (H1) and Livingston (L1) detectors. In this paper, we extend our search to cover the region in parameter space in which the signal response is very high in one detector, but small in the other (this regime is shown as the teal region[1] in Fig. 1. It is in general challenging to reliably compute the false alarm rate (FAR) of a trigger in this region, because throughout the entire observing run, there are only a small number of triggers that (a) have comparably high SNRs, (b) are well fit by similar waveforms, and (c) pass our vetoes. The fact that we cannot realistically simulate interferometer data prevents us from empirically measuring the FAR.

We could empirically measure the number of fainter triggers and extrapolate the distribution. However, loud triggers are mainly produced due to anomalous detector behavior, i.e., so-called glitches [6,7], rather than stationary Gaussian random noise. Since we do not completely understand the physical or instrumental origins of glitches [8], extrapolations of their distributions to higher values of SNR come with significant uncertainty.

In our previous analysis in Ref. [2], we did not search for events with highly incommensurate SNRs in the two LIGO detectors (the region of phase space with $\rho_{\rm L}^2 > 66$ and $\rho_{\rm H}^2 < 16$, where ρ is the SNR and the subscript refers to the detector), since we imposed a threshold on the SNR for collecting triggers (this threshold equaled 4 in the banks covering massive BBH mergers). Given the higher sensitivity of L1 compared to H1 on average throughout O2, we estimated that our cut on single-detector SNR reduced the

---

[*]bzackay@ias.edu
[1]The analogous region with high SNR in H1 corresponds to much smaller sensitive volume.
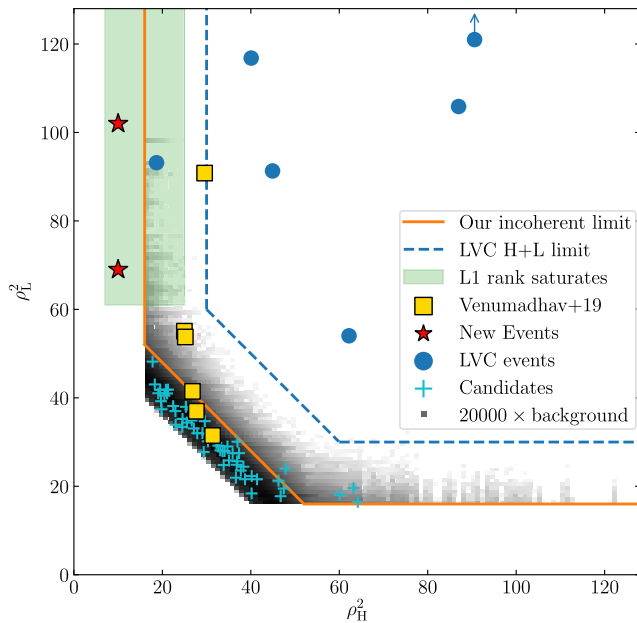
FIG. 1. Incoherent Hanford (H1) and Livingston (L1) SNR^2 for coincident and background triggers (computed using 20 000 time slides), for all the sub-banks with events. The blue and orange lines are approximate incoherent detection limits for analyses in Refs. [1,2], respectively, restricted to using H1 and L1 data only. GW170814 has $\rho_L^2 = 170$, higher than shown here (indicated with an arrow), and GW170608 is not shown because its H1 data is not part of the bulk O2 data release. Figure adapted from Ref. [2].

sensitive volume of our search by ∼10%. Note that even when the two LIGO detectors have comparable sensitivity, astrophysical sources with unfortunate sky positions and orbital orientations can produce signals of different strengths in the detectors. The search described in this paper revealed two additional interesting BBH triggers that are above the thresholds of significance for being called events [1].

One event, GW170817A (not to be confused with the binary neutron star merger event GW170817 [9]) comes from a pair of black holes with a very high total mass ∼100 $M_\odot$. The existence of such a BBH system is informative about theories of the evolution and death of massive stars, which generically predict an upper mass gap at ∼50 $M_\odot$ for stellar mass black holes.

Efforts to estimate the parameters of the other candidate, GWC170402, yield significant evidence that the signal is not fully described by waveforms of the dominant harmonic mode for circular binaries with aligned spins. This will be discussed thoroughly in a forthcoming paper. At the time of writing, we do not have a waveform model that completely models the observed signal, and hence we do not attach a "GW" prefix but a temporary GWC for "GW candidate."

The paper is organized in the following way: In Sec. II we discuss the technique and its application to the O2 data.

In Sec. III we summarize the analysis results. In Sec. IV we present the results of parameter estimation for GW170817A, and discuss its relevance to astrophysical formation scenarios for massive black holes. We finish with our conclusions in Sec. V, and provide some extra details in the Appendices.

II. METHODOLOGY AND RESULTS

In this section, we describe the methods we use to analyze and assign significance to single detector triggers, and present results alongside. We begin with a brief overview below, and expand upon the details in individual sections.

In this analysis, we focus on triggers from template banks for relatively massive BBH mergers, with detector-frame chirp masses ≥20 $M_\odot$. This part of parameter space is most promising for the search presented in this paper, because it contains twelve of the BBH mergers that have been detected as coincident H1 and L1 triggers so far, and hence there is a significant chance that one or more events from similar sources may have been missed by previous analyses (due to SNR cuts we imposed when collecting triggers, or approximations used for the coherent score that were valid in the high SNR regime).

To perform this specialized search, we first define a set of significant L1 triggers, which are so loud that it is extremely unlikely that Gaussian random noise produces them, even over the entire length of the O2 run. However, glitches can produce such loud triggers, and hence we rank these triggers not by their SNR, but by an empirical measure of how frequently known glitches contaminate their surrounding phase space. Section II A presents details of this ranking and justification for it.

We then examine the strain data from the less-sensitive detectors [H1 and/or V1 (i.e., Virgo)] for counterpart signals of each of the above L1 triggers. We define a score that, given a signal in L1, coherently computes the joint likelihood from the data in all available detectors, and marginalizes over extrinsic parameters of the source. We derive this score and its properties in Sec. II B.

We next need to combine the information from the more- and less-sensitive detectors (L1, and H1 and/or V1, respectively) and estimate a final false alarm rate (FAR) for the triggers. We describe our method to do so in Sec. II C.

The FAR quantifies the rate at which detector noise produces triggers above a threshold. In a similar manner to searches of coincident triggers, we need to compare this rate to the rate at which the known astrophysical population of mergers would produce the triggers, and estimate the probability of astrophysical origin (p_{astro}) for the candidates. Section II D outlines our method to accomplish this.

Finally, in Sec. II E, we validate our methods by applying them to the event GW170818, which is a highly significant GW event in the official catalog released by the LVC [1]

that lies in the region of phase-space covered by this search (the blue circle within the teal region in Fig. 1).

A. Ranking single detector triggers

The L1 detector was more sensitive over most of the O2 run, and hence we expect that loud single detector events in H1 (with $\rho_L^2 < 16$) are much rarer than similar events in L1. Moreover, we empirically observe that the L1 detector produces a much lower number of loud triggers that pass our signal-quality vetoes (i.e., glitches). Hence, we focus our efforts toward characterizing loud L1 triggers.

Our previous search within coincident triggers used rank functions to sort triggers from the two LIGO detectors by their significance [2]. Rank functions empirically quantify the probability that the underlying noise process produces triggers at a given value of SNR; we computed them separately for each detector, and for different regions of the source parameter space. In particular, our search used several template banks (logarithmically spaced in chirp mass), each in turn divided into sub-banks that captured the variety of waveform amplitude profiles [10]. We computed rank functions separately for each sub-bank, since the non-Gaussian tails of the single-detector trigger distribution varied significantly as a function of parameters. This allowed us to assess the significance of coincident triggers by consistently and locally estimating the effects of glitches.

In Ref. [2], we noted that the rank functions empirically followed their behavior in the Gaussian-noise case to higher values of SNR in those sub-banks in which we found real events. It is especially remarkable that the sub-bank BBH (3,0) was essentially clean (i.e., without glitches, see Appendix A for more details); the five loudest L1 triggers in this sub-bank belonged to GW events that were confirmed using coincident H1 triggers. Further investigations show that there are dramatic inhomogeneities in the rates at which templates produce triggers that pass our vetoes (i.e., some templates disproportionately trigger on glitches, relative to the bulk). Appendix A presents evidence for this phenomenon.

This is a natural outcome if there is some finite number of "glitch waveforms," in which case only templates that are similar enough to these waveforms produce loud veto-passing triggers (for previous work that reached similar conclusions, see Refs. [11,12]). Guided by this intuition, we identify "glitch-prone" templates using the following empirical procedure:

(1) Collect all "triggers of interest," defined as L1 triggers with $\rho_L^2 > 66$ that pass our vetoes, with the best fit waveform having a chirp-mass $m_c > 20\,M_\odot$, and record the template with the highest value of ρ_L^2 for each trigger. We chose the bound on ρ_L^2 such that random Gaussian noise would produce (in expectation) only one trigger like this over the entire run: we computed it using the survival function of a chi-squared distribution with five degrees of freedom[2] (amplitude, phase, time, mass, and spin), 10^2 independent templates, and 118 days of data. The Gaussian noise hypothesis is unlikely for triggers above this bar, and the remaining explanations are that they are either glitches or genuine signals.

(2) Define as suspected L1 glitches all triggers that pass our vetoes and are not already detected GW events, have $\rho_L^2 > 55$ and have available H1 data. We computed this bound in the same way as before, but with one independent template (hence there will be a few Gaussian noise candidates in here, but in practice, glitches dominate this distribution).

(3) For each trigger of interest, count the number of suspected glitches whose templates have a significant match (≥ 0.9) with that of the trigger. We use this as an effective measure of the impact of glitches in the associated region of phase-space (note that this implicitly assumes that each template accounts for an equal volume of phase space, which is the prior we adopted in our previous analysis [2]).

We then rank the triggers of interest according to (a) the number of suspected L1 glitches, and then (b) the L1 SNR, ρ_L. We do not assign a higher weight to ρ_L in the ranking, since the number of glitches does not steeply decline as a function of SNR (in particular, it does not exhibit the exponential tails characteristic of chi-squared variables). While this procedure and the chosen thresholds are well motivated, it is important to verify that the chosen parameters do not sensitively affect the inference down the line, keeping it as objective as possible. We therefore repeated this procedure with slight threshold changes, and confirmed that the ranking is invariant to this choice (as long as the similar trigger threshold is high enough to reject the Gaussian contribution).

Since the Gaussian noise hypothesis is not viable for the triggers of interest, and we accounted for the effect of glitches in a conservative way (i.e., without over-interpreting high values of ρ_L), this ranking represents our best degree of belief in a L1 trigger being of astrophysical origin (before considering other detectors, or previous detections).

Table I gives the results of this ranking procedure applied to the triggers of interest.

---

[2]The parameters optimized in the search are amplitude, phase, time, c_2 and c_3 (and sometimes c_4, in the relevant banks). c_2 and c_3 are scalar parameters that define the internal bank coordinates we are using, see [10] for more details. When projecting real waveforms onto the bank coordinates, chirp-mass and effective spin changes influence the (c_2, c_3) coordinates the most (though sometimes nonlinearly) the effects of mass ratio are also non-negligible. The counting of degrees of freedom is qualitative, and is expected to explain the trigger abundance only up to one order of magnitude error.

TABLE I.   Triggers ranked solely based on data from the Livingston (L1) detector. The ranking is based on the number of similar triggers with L1 $SNR^2 = \rho_L^2 > 55$ that pass our vetoes, which assesses the relative tendency of glitches in L1 to produce similar spurious background triggers. Note that this simple ranking marks essentially all previously confirmed loud ($\rho_L^2 > 60$) BBH mergers based on the L1 triggers alone. The next three columns quantify the evidence for the astrophysical nature of the triggers from data in the Hanford (H1) detector, in terms of our coherent score $\mathcal{S}$ [see Eq. (1)]: $C(\mathcal{S}|H_0)$ ($C(\mathcal{S}|H_1)$) is the probability of obtaining a coherent score higher (lower) than that of the trigger in a random segment of H1 data without a signal (with an injected signal with consistent intrinsic parameters). Note that the new triggers (marked bold) that have high ranks based on L1, also have significantly low values of false alarm probability, $C(\mathcal{S}|H_0)$. The triggers ranked 9–11 are judged as background using the data in this table alone (that is, using the H1/V1 data). Both their $C(\mathcal{S}|H_0)$ is too high, indicating a random location at H1/V1, and $C(\mathcal{S}|H_1)$ is tiny (indicating that we should have seen a stronger counterpart had they been real events).

L1 rank	GPS time	$\rho_L^{2\,\text{a}}$	# similar triggers	$C(\mathcal{S}|H_0)$	$C(\mathcal{S}|H_1)$	$\frac{P(\mathcal{S}|H_1)}{P(\mathcal{S}|H_0)}$	Comment
1	1187058327.068	93.1	0	$<10^{-3}$	0.16	37	GW170818[b]
2	1187529256.504	92.1	0	$\cdots$	$\cdots$	$\cdots$	GW170823
3	1169069154.564	90.8	0	$\cdots$	$\cdots$	$\cdots$	GW170121
4	**1175205128.565**	**72.9**	**0**	**0.015**	**0.022**	**0.547**	**GWC170402**
5	1186741861.51	174.6	1	$\cdots$	$\cdots$	$\cdots$	GW170814
6	1167559936.584	107.3	1	$\cdots$	$\cdots$	$\cdots$	GW170104
7	1186302519.731	118.6	2	$\cdots$	$\cdots$	$\cdots$	GW170809
8	**1186974184.716**	**98.5**	**5**	**0.028**	**0.055**	**0.98**	**GW170817A**
9	1174043898.842	75.7	9	0.36	0.001	0.008	Background
10	1170885005.109	66.4	16	0.49	0.003	0.013	Background
11	1178083239.592	74.4	22	0.34	0.003	0.016	Background
Removed[c]	1173477193.704	69.2	1	0.38	0.014	0.011	Artifacts present

[a]We report only the relevant integrals of the S values as the ρ_H^2 for these events is not well defined (the integral in Eq. (1) may not be dominated by the maximum likelihood region).

[b]For the purpose of demonstrating our new methodology, we present numbers corresponding to analyzing data from only the two LIGO detectors, even though Virgo detected GW170818 at $SNR \simeq 4$[1].

[c]We removed this candidate as its Livingston spectrogram shows immediately obvious signs of nonstationary activity, or "glitchy" behavior (see Fig. 9 in Appendix B). We include it in the list for completeness.

B. Coherent score from less-sensitive detectors

The procedure of Sec. II A relies only on the L1 data for a given trigger (note that we use the absence of H1 triggers to build a list of suspected glitches, so the procedure as a whole requires data from Hanford). For every L1 trigger of interest, we now search for weak counterpart GW signals in other detectors whenever coincident data is available.

For CBC sources, counterpart signals should agree with the L1 signal in terms of the shape of the waveform, but in general differ in the arrival time, the amplitude normalization, and in the phase constant (this is strictly true only for the dominant (2, 2) harmonic of the GW signal). These are determined by extrinsic parameters, which we denote by the symbol $\boldsymbol{\Theta}_{\text{ext}}$.

Even when the two LIGO detectors have similar sensitivities, since they are not perfectly antialigned, astrophysical sources at special sky locations and with special inclinations can produce disparate SNRs in H1 and L1. However, such source configurations are fine-tuned and are thus *a priori* disfavored. For this reason, it is necessary to marginalize over the possible values of the extrinsic parameters. For this purpose, we use a conditional coherent score (hereafter coherent score for brevity) $\mathcal{S}$ that we can efficiently compute for each trigger.[3]

First, we fix the intrinsic CBC parameters, $\boldsymbol{\Theta}_{\text{intr}}$, (detector-frame masses, spins) to their best-fit values from the L1 data alone. The coherent score, $\mathcal{S}$, is the logarithm of the Bayesian evidence for a joint fit to the L1 and H1 data (we can also include V1 data when available), marginalized over all possible combinations of extrinsic parameters $\boldsymbol{\Theta}_{\text{ext}}$:

$$e^{\mathcal{S}} := \int \mathcal{D}\boldsymbol{\Theta}_{\text{ext}} \Pi(\boldsymbol{\Theta}_{\text{ext}}) \mathcal{L}(\boldsymbol{d}|\boldsymbol{\Theta}_{\text{intr}}, \boldsymbol{\Theta}_{\text{ext}}). \qquad (1)$$

In the above equation, the symbol $\Pi(\boldsymbol{\Theta}_{\text{ext}})$ denotes the properly normalized prior for all 7 extrinsic parameters: sky position RA and DEC, polarization angle ψ, orbital inclination ι, orbital phase φ, geocentric arrival time t_c, and the source luminosity distance d_L. The quantity $\mathcal{L}(\boldsymbol{d}|\boldsymbol{\Theta}_{\text{intr}}, \boldsymbol{\Theta}_{\text{ext}})$ is the likelihood function, which is given by

---

[3]Note that this is different from, but analogous to, the coherent score we applied to two-detector coincident triggers in our previous joint search of H1 and L1 triggers, which is an approximation that works in the high SNR limit [2].
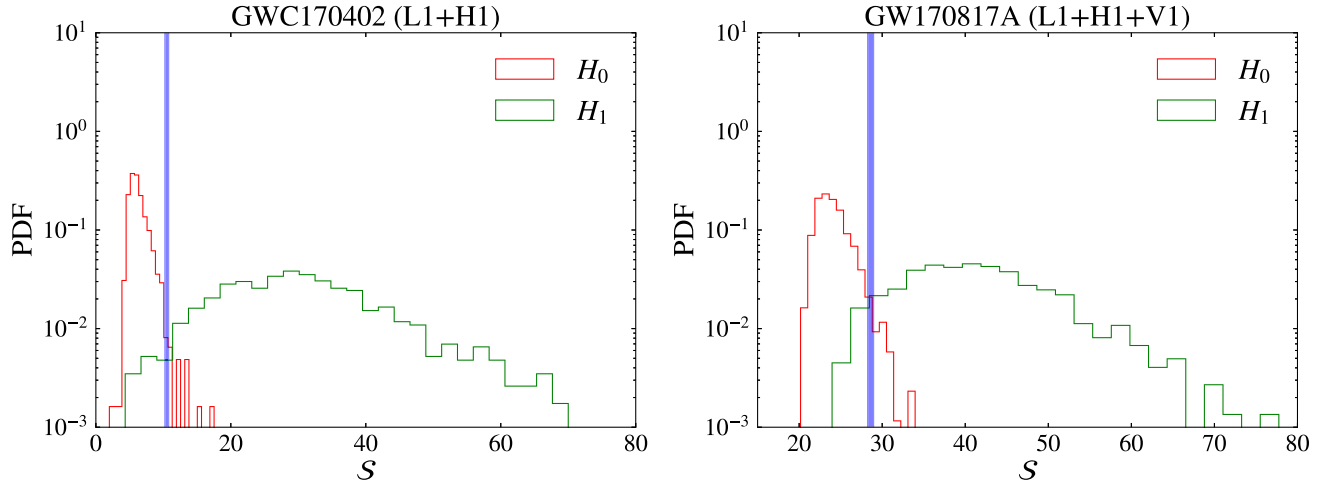
FIG. 2. Distributions of coherent score S for the top two candidates in Table I using data from the less sensitive detectors (the left and right panels, respectively, show scores for GWC170402 at GPS time 1175205128.565 from H1, and GW170817A at GPS time 1186974148.716 from H1 and V1; see the note in the text about H1 data for GW170817A). The symbols H_0 and H_1 indicate the noise and astrophysical hypotheses, respectively, under which we derive distributions using 1000 time slides in each case. The vertical shaded region indicates the spread in the values of S for different choices of the fiducial intrinsic parameters Θ_{intr} consistent with L1 data.

$$\ln \mathcal{L}(\boldsymbol{d}|\boldsymbol{\Theta}_{\text{intr}}, \boldsymbol{\Theta}_{\text{ext}}) = \sum_i \left[\langle d_i | h_i(\boldsymbol{\Theta}_{\text{intr}}, \boldsymbol{\Theta}_{\text{ext}}) \rangle \right.$$
$$\left. - \frac{1}{2} \langle h_i(\boldsymbol{\Theta}_{\text{intr}}, \boldsymbol{\Theta}_{\text{ext}}) | h_i(\boldsymbol{\Theta}_{\text{intr}}, \boldsymbol{\Theta}_{\text{ext}}) \rangle \right].$$

$$(2)$$

Here, the index i runs over the different detectors, and d_i and $h_i(\boldsymbol{\Theta}_{\text{intr}}, \boldsymbol{\Theta}_{\text{ext}})$ are the strain data and the signal in the ith detector, respectively. We use $\langle \cdots | \cdots \rangle$ to denote the standard matched filter overlap.

For the priors $\Pi(\boldsymbol{\Theta}_{\text{ext}})$, we use a uniform prior distribution on the arrival time t_c, and isotropic priors for the source position on the sky and the orbital orientation of the binary. For the luminosity distance d_L, we assume a constant volumetric density in Euclidean space within $0 < d_L < 10$ Gpc. In practice, we analytically marginalize over d_L and φ.

Note that the most rigorous definition of the coherent score [Eq. (1)] should marginalize over both intrinsic ($\boldsymbol{\Theta}_{\text{intr}}$) and extrinsic ($\boldsymbol{\Theta}_{\text{ext}}$) parameters, instead of fixing the former to their best-fit values from the L1 data. Since the full parameter space is high-dimensional, this significantly increases the computational cost of evaluating S. Under the signal hypothesis, the L1 SNR ρ_L is high enough to constrain the intrinsic parameters as well as they are in a joint fit to L1 and H1 (and V1 if available) data. The values of individual intrinsic parameters (such as masses and spins) are often substantially correlated, but since the degenerate combinations map to nearly the same waveform, and intrinsic parameters are largely uncorrelated with extrinsic ones, it is safe to use the best-fit combination. To verify this, we computed scores for several different

choices of intrinsic parameters that achieve a squared SNR in L1 that is up to five points lower then the best achieved. The obtained S values are plotted as blue vertical lines in Fig. 2. Since $P(S|H_0)$ and $P(S|H_1)$ are template independent, and that different templates consistent with the observed L1 data produce practically the same S score, all the reported values in Table I are robustly measured.

Given a trigger in the L1 detector, to interpret its associated coherent score S, we need to consider its expected probability density function (PDF) under two competing hypotheses:

(1) *Astrophysical hypothesis* (H_1): the L1 trigger is caused by an astrophysical gravitational wave signal, and hence it must have consistent counterpart signals in the other detectors. Under this hypothesis, the coherent score has an expected PDF $P(S|H_1)$.

(2) *Noise hypothesis* (H_0): the L1 trigger is caused by noise processes in the detector. In this case, there should not be any correlated counterpart signals in the other detectors. Under this hypothesis, the coherent score has an expected PDF $P(S|H_0)$.

These distributions inform us about the significance of the event in two ways. Firstly, if the L1 single detector trigger is due to an astrophysical event, we expect the coherent score S to be more consistent with the distribution $P(S|H_1)$ than with $P(S|H_0)$. Secondly, under the noise hypothesis H_0, the probability for S to be greater than the measured value is analogous to the FAR computed in coincidence analyses [2].

We determine both the distributions ($P(S|H_1)$ and $P(S|H_0)$) locally and independently for each L1 trigger, by empirically sampling from them. The local measurement ensures that $P(S|H_0)$ is representative of each trigger since the noise background in H1 can fluctuate significantly over time. More importantly, the relative sensitivities

between different detectors vary substantially over the run, which affects the relative strengths of any astrophysical signals within, and hence dramatically changes $P(\mathcal{S}|H_1)$ from one trigger to another.[4] We can determine the distributions locally and empirically without any extrapolation because we only need to sample tail probabilities $\geq 10^{-3}$.

We determine $P(\mathcal{S}|H_0)$ by keeping the L1 strain series fixed, but sliding the strain series in the other detector(s) in time by more than two seconds, which well exceeds the physically allowed time delay relative to L1. We then evaluate the coherent score $\mathcal{S}$ as defined previously at this unphysical lag. In practice, we restrict the extra time lag to be within a few thousand seconds to obtain a local estimate. Our search pipeline also flags ill-behaved segments of data during its preprocessing phase, and masks and in-paints these segments (as well as segments marked by the LVC's quality flags) to avoid contaminating neighboring seconds [3,13]; we exclude these segments from the time slides. We repeat this procedure a large number of times and generate samples from $P(\mathcal{S}|H_0)$.

We determine the distribution $P(\mathcal{S}|H_1)$ using a similar procedure as above, with the difference being that we inject counterpart gravitational wave signals into the strain data in the other detector(s) after applying time slides. We generate injections by fixing the intrinsic CBC parameters $\mathbf{\Theta}_{\text{intr}}$ to their best-fit values, and generating extrinsic parameters $\mathbf{\Theta}_{\text{ext}}$ from their posterior distributions (both obtained using only the L1 data). We estimate the distribution $P(\mathcal{S}|H_1)$ by repeating the above procedure several times.

Figure 2 shows the distributions $P(\mathcal{S}|H_0)$ and $P(\mathcal{S}|H_1)$ for the top two triggers in Table I that are not already confirmed events. Note that for GW170817A, the spectrogram of the H1 data (i.e., the less-sensitive detector, which was not used to identify the trigger) at the time of the event shows artifacts that are localized to a few bands in the frequency domain. Hence, before analyzing the H1 data, we removed frequencies between 68–73 Hz, and 92–96 Hz by applying notch filters (implemented as Bessel filters with critical frequencies at the edges of the quoted frequency intervals).

C. Determining the false-alarm rate

In this section, we describe our procedure for assigning false-alarm rates to the L1 triggers using the information from the other available detectors. We first compute the false alarm probability (FAP) for a trigger of interest, indexed by i, given its coherent score $\mathcal{S}_i$ (which is based on the data in H1 and/or V1, conditioned on the loud trigger in L1), as $\text{FAP}(i) = P(\mathcal{S} > \mathcal{S}_i|H_0) = C(\mathcal{S}_i|H_0)$. This is the survival function for the coherent score under the noise

hypothesis, H_0. In order to obtain the false alarm rate, we need to combine this FAP with the occurrence rate of the L1 trigger, and the effective look elsewhere effect.

The triggers of interest were all chosen such that their scores are well above the thresholds for being produced in Gaussian noise. As we mentioned in Sec. II A, we would like to avoid overinterpreting high values of SNR in L1, since extrapolations of the distribution from lower values of SNR are unreliable. Hence, we limit the information from L1 to the rank of the trigger in the list of triggers of interest (this skews to being conservative in interpreting SNR, since the occurrence rates of triggers with a given rank are bounded below by 1 per O2 observing run, and penalizes triggers from regions of parameter space that are affected by glitches).

The rate of triggers being in the first place in the L1 ranking is 1 per O2 (by definition), and hence the probability of the first-place trigger having a FAP $< \epsilon$ based on the other detectors is ϵ per O2. The first trigger on our list has a FAP of 0.015, and hence its false alarm rate is

$$\text{FAR}^{-1}_{\text{GWC170402}} = 60 \text{ O2} \approx 19 \text{ yr} \qquad (3)$$

Based on this false alarm rate alone, the candidate is well above the threshold significance to be considered interesting [1]. If this trigger also has a high probability of being astrophysical in nature (see more details in Sec. II D), we can add it to the catalog of events, in which case the trigger ranked second becomes the new top candidate. The second trigger on the list has a FAP of 0.028, and by a similar argument as above, a FAR of:

$$\text{FAR}^{-1}_{\text{GW170817A}} = 36 \text{ O2} \approx 11.5 \text{ yr.} \qquad (4)$$

This is also well above the threshold of significance to be considered interesting; we will estimate a value of p_{astro} for this trigger in Sec. II D.

The triggers further down the list do not have compelling evidence from H1/V1 (see the $C(\mathcal{S}|H_1)$ column in Table I), and hence we terminate the procedure at this point. Note that since the two events are at the top of the list, we effectively have no significant corrections due to the look elsewhere effect. If events occur further down the list, we would need a more involved procedure that carefully takes the look elsewhere effect into account when estimating their significance.

Note also that there was one candidate removed from the search in L1 in a nonautomatic fashion, because of the poor L1 data quality around the candidate (see Appendix B), some readers may therefore want to adopt slightly more conservative values for the false alarm rate: $\text{FAR}^{-1} = (1/(30\text{O2})$ for GW170402 and $\text{FAR}^{-1} = 1/(18\text{O2})$ for GW170817A).

---

[4]The same effect is operative in coincidence analyses as well. We accounted for it in Refs. [2,3] using an approximation that is valid in the limit of high SNR in H1.

D. Determining the probability that a trigger is of astrophysical origin

Apart from the FAR, searches in coincident triggers also report a probability of astrophysical origin (p_{astro}) for the candidates. Given a trigger with a set of properties $\mathcal{T}$, the definition of p_{astro} is

$$p_{astro} = \frac{R(\mathcal{T}|H_1)}{R(\mathcal{T}|H_0) + R(\mathcal{T}|H_1)} = \frac{\mathcal{Q}}{1 + \mathcal{Q}}, \quad (5)$$

with $\mathcal{Q} = R(\mathcal{T}|H_1)/R(\mathcal{T}|H_0)$. In order to obtain an estimate of p_{astro} for the candidates in this paper, we would need to empirically measure the distributions of SNR for triggers for similar templates in Livingston, and extrapolate them to higher values (the candidates we are discussing are the loudest triggers in their distributions, see Fig. 3). As we mentioned in the introduction, we do not have a physical model for glitches, due to which the results of this extrapolation are subject to significant uncertainties.

In a desire not to overinterpret the high values of ρ_L^2, we conservatively restrict the information from L1 to the fact that $\rho_L^2 > 66$, and that the corresponding templates are in the "clean," or "glitch-free" region of parameter space. Note that all previously discovered black hole mergers are comfortably inside the 'clean' region of parameter space.

triggers similar to GW170817A

triggers similar to GWC170402

ρ_L^2
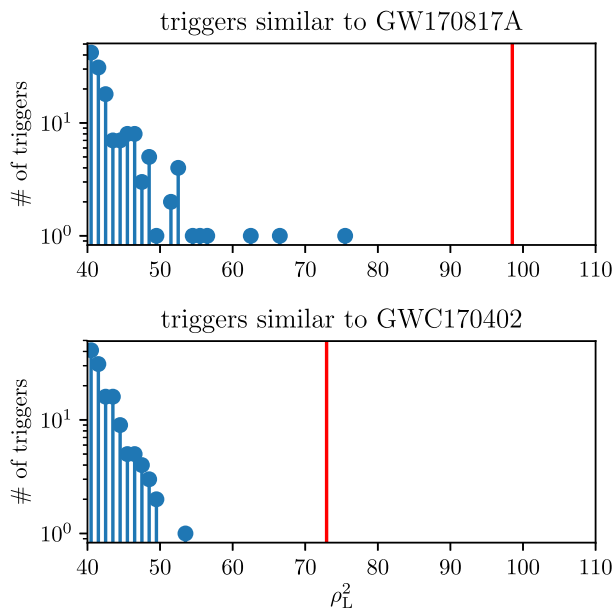
FIG. 3. Distributions of L1 SNR^2 for triggers for templates that are similar (match > 0.9) to the best-fit templates for the two newly found events, that occur at times when the H1 detector is operative. Vertical red lines mark the values of ρ_L^2 for the two events. To give context to the amount of phase space that is included in this plot, the upper (lower) panel includes triggers from 28% (0%) of bank BBH 4, and 1.8% (3.6%) of bank BBH 3.

We then proceed with the calculation of the rate ratio,

$$\mathcal{Q} = \frac{R(\mathcal{T}|H_1)}{R(\mathcal{T}|H_0)} = \frac{R(\rho_L^2 > 66, \text{clean}|H_1)}{R(\rho_L^2 > 66, \text{clean}|H_0)} \frac{P(\mathcal{S}|H_1)}{P(\mathcal{S}|H_0)}, \quad (6)$$

where the first and second terms in both the numerator and denominator give the information coming from the stronger detector and that from the weaker detectors. The rate

$$R(\rho_L^2 > 66, \text{clean}|H_0) \quad (7)$$

effectively represents the rate of triggers that are above the trigger in question (including itself) in the Livingston based ranking. Since these two triggers are at the top of the list, we can place an upper limit of once per O2 on this rate.[5] We caution the reader that this estimate is very uncertain, because we cannot run the experiment of making such a rank on fake data many times (as we cannot simulate real single-detector data). We should consider this number as being subject to order unity uncertainty, since, in principle, there could have been a glitch above our events in the rank.

$$R(\rho_L^2 > 66, \text{clean}|H_1)$$
$$= \mathcal{R}_{O2} \times \frac{P(\rho_L^2 > 66)}{P(\rho_H^2 + \rho_L^2 > \rho_{threshold}^2, \rho_{H,L}^2 > 16)}, \quad (8)$$

where $\mathcal{R}_{O2} = 13$ is the total reported rate for events in banks BBH 3 and BBH 4, as reported in [2], and we account for the different volumes that the two analyses are sensitive to. We use the combined rate because the analysis in this paper covers the union of these two banks (as we mentioned at the beginning of Sec. II). The second term on the right-hand side of Eq. (8) is the ratio of sensitive volumes for the coincidence and single-detector analyses, and depends on the relative sensitivities of the detectors in the network (we only consider H1 and L1 when estimating significance). This volume ratio evaluates to 0.33 (0.5) when the detectors are equally sensitive (L1 is 40% more sensitive than H1) as is relevant to the case of GWC170402 (GW170817A).

Substituting all the above factors, the rate ratios for the two events evaluate to:

$$\mathcal{Q}_{GWC170402} \approx 2.15$$
$$\mathcal{Q}_{GW170817A} \approx 6.37, \quad (9)$$

which implies that their probabilities of being of astrophysical origin are

---

[5]The p_{astro} values do not change upon inclusion of the removed trigger (see Appendix B), as the 1/O2 values used was previously an upper limit, and now is the actual measured glitch rate.
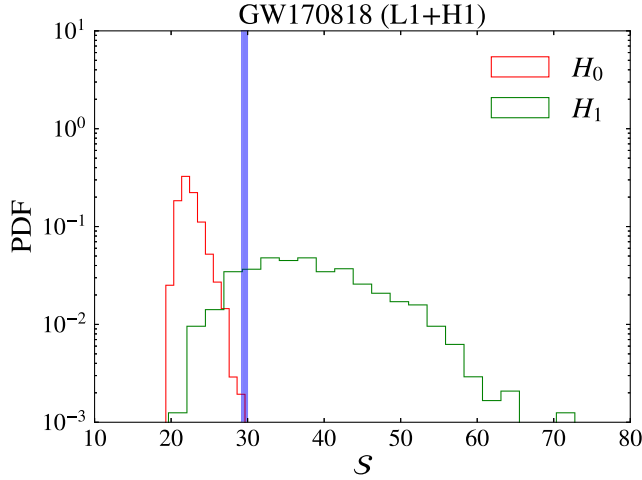
FIG. 4. Demonstration of our coherent score $\mathcal{S}$ with the LVC event GW170818 using data from only the two LIGO detectors. The notation, and the number of time slides used, are identical to those of Fig. 2.

$$p_{\text{astro}}(\text{GWC170402}) = 0.68$$
$$p_{\text{astro}}(\text{GW170817A}) = 0.86. \tag{10}$$

Note that this does not include any information from the fact that these triggers are outliers in their respective, locally estimated background distributions. If we had a credible way of accounting for this fact, it would only

increase the inferred values of p_{astro}. Figure 3 shows the local background distributions for triggers for templates similar to those for the two events.

E. Validation using GW170818

We also validate the above procedure by applying it to GW170818, a BBH event previously reported by the LVC [1]. It was marked as an L1 single detector trigger by the `PyCBC` pipeline but was not considered for coincidence analysis because its SNRs at H1 and V1 were below the threshold for collection. It was initially identified by the `GstLAL` pipeline as a L1-V1 double detector trigger until a H1 counterpart signal was later confirmed in the offline search. It was also confirmed by a refined analysis with the `PyCBC` pipeline [14].

GW170818 is an ideal example to demonstrate how the astrophysical nature of a single detector trigger can be validated by evaluating the coherent score, as it has a high SNR at L1 ($\sim$10) but very low SNRs at H1 and V1 (both $\sim$4). Figure 4 shows the coherent score $\mathcal{S}$ for this event using the data in L1 and H1 alone, and its distributions under the noise and astrophysical hypotheses. The high value for the coherent score, $\mathcal{S} = 29.7$, measured from the consistency of the data in the two LIGO detectors, is an obvious outlier relative to $P(\mathcal{S}|H_0)$, with none of our 1000 Montecarlo realizations yielding a higher value for the score. In contrast, it is fully consistent with typical values drawn from $P(\mathcal{S}|H_1)$. At $\mathcal{S} = 29.7$, the probability density
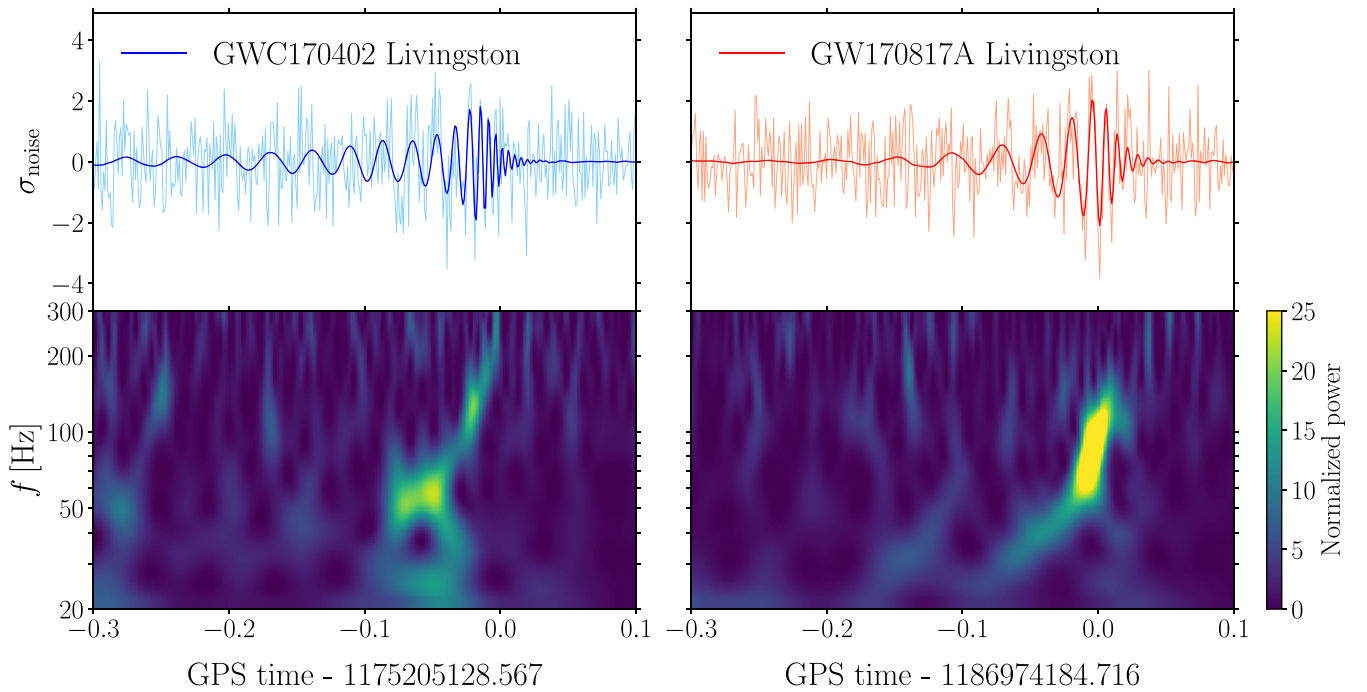

FIG. 5. Two BBH candidates GWC170402 and GW170817A initially selected as significant L1 single detector triggers. Upper panels show the whitened strain series around the trigger times (light colored curves), with the network maximum likelihood `IMRPhenomD` waveforms overplotted (dark colored curves). The corresponding spectrograms are shown in the lower panels.

for the coherent score under the astrophysical hypothesis is more than 30 times higher than that under the noise hypothesis.

In Table I, this is the trigger with the highest L1 SNR among those that have not been validated in the two-detector joint analysis, and it has no similar glitches. Considering these facts, we are able to assign an inverse FAR better than 1000 O2 for GW170818 purely from the coherence of the recovered signals in L1 and H1. Hence, we are able to confirm its astrophysical origin even without confirmation from the Virgo detector.

III. SUMMARY OF RESULTS

Table I summarizes our results on O2. Among the eight highest (L1-based) ranking events, six were previously detected BBH events (that were detected using Hanford (H1) and Virgo (V1) data), which gives us some confidence in the ability of the ranking statistic to identify interesting triggers.

For the other two triggers, we detect faint counterpart signals in Hanford with a locally measured inverse false-alarm rate of $FAR^{-1} > 36\,O2 \approx 11.5$ yr. Note that for GW170817A, the H1 data needs to be cleaned of artifacts in the frequency domain by notching out frequencies 68–73 Hz, and 92–96 Hz. The excess noise in these bands was noted in visual inspection. It appeared abruptly and was elevated for an extended period of time.

We tested triggers further down on the list, and do not find any significant supporting evidence from the Hanford detector for any of them. We additionally verified that the distribution of the computed scores (quantifying multi-detector coherence) for these fainter triggers is consistent with the predicted one with no signal (and hence consistent with pure background events in the Livingston detector).

We name the two new events GWC170402 and GW170817A according to the date of occurrence. Figure 5 shows the spectrograms of the strain data in the Livingston detector around the two events.

IV. ASTROPHYSICAL IMPLICATIONS OF GW170817A

We perform parameter estimation for GW170817A by using relative binning [15] to compute the likelihood, and the PyMultiNest code [16] to generate samples from the posterior. As noted earlier, we had to apply notch filters to the H1 data to determine the significance of this event; we apply the same filters before performing parameter estimation as well. Figure 6 presents posteriors for the source-frame masses, effective spin, and the redshift, marginalized over other parameters.

The inferred source frame total mass of GW170817A is $M_{tot}^{src} = 98_{-11}^{+17}\,M_{\odot}$, while its inferred effective spin is $\chi_{eff} = 0.5_{-0.2}^{+0.2}$. It is interesting to consider the individual component masses: the primary black hole has a
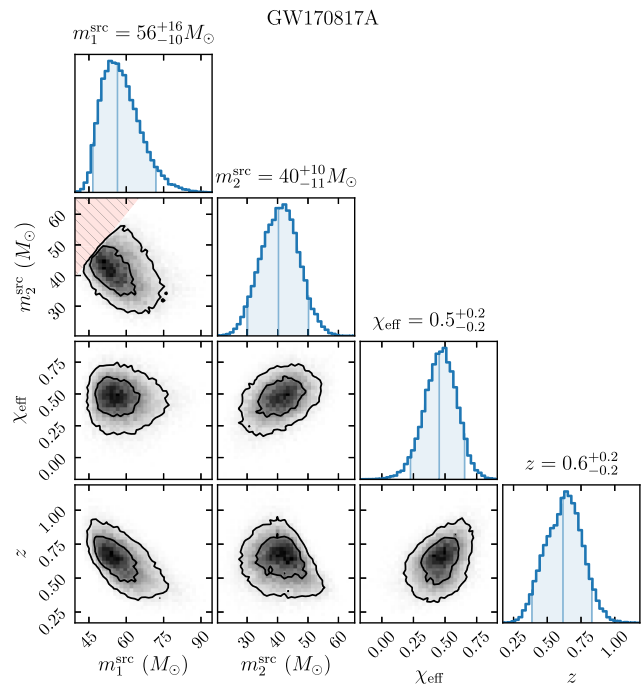

FIG. 6. Marginalized posteriors for GW170817A. Two-dimensional contours enclose 50% and 90% of the distribution. In the one-dimensional posteriors, vertical lines mark the 0.05, 0.5 and 0.95 quantiles. We compute likelihoods using the IMRPhenomD [17] waveform model, and adopt a prior that is uniform in detector-frame m_1, m_2, χ_{eff} and luminosity volume.

source-frame mass of $m_1^{src} = 56_{-10}^{+16}\,M_{\odot}$ (the limits indicate 95% confidence intervals), which would put it at the heaviest end of the merging black holes discovered so far (see Ref. [1]).

The presence of such massive individual black holes is potentially informative about the formation of stellar mass black holes from massive progenitor stars at the end of their lives. Models of stellar evolution predict that extremely massive stars with Helium core masses in the range of ~50–130 $M_{\odot}$ do not produce BH remnants with these masses, since they either explode as pair instability super-novae and leave no remnants, or shed substantial mass via the pulsational pair instability and leave lower-mass remnants [18]. However, the location of this mass gap is subject to large uncertainties, since it depends on the phenomenology of mass loss [19]. Extremely metal-poor stars may collapse into BH remnants even as massive as 70–80 $M_{\odot}$ [18,20]. It has also been suggested that dense stellar systems can harbor massive BBHs in which one (or both) of the components is a product of a prior merger, which would produce BHs above any mass gap [21].

Previous work has used the LVC detections to jointly infer the properties of the population of BBHs in the Universe [22–26]. Partially motivated by the above astrophysical considerations, some of the models considered have an upper cutoff to the mass of the merging BHs: the
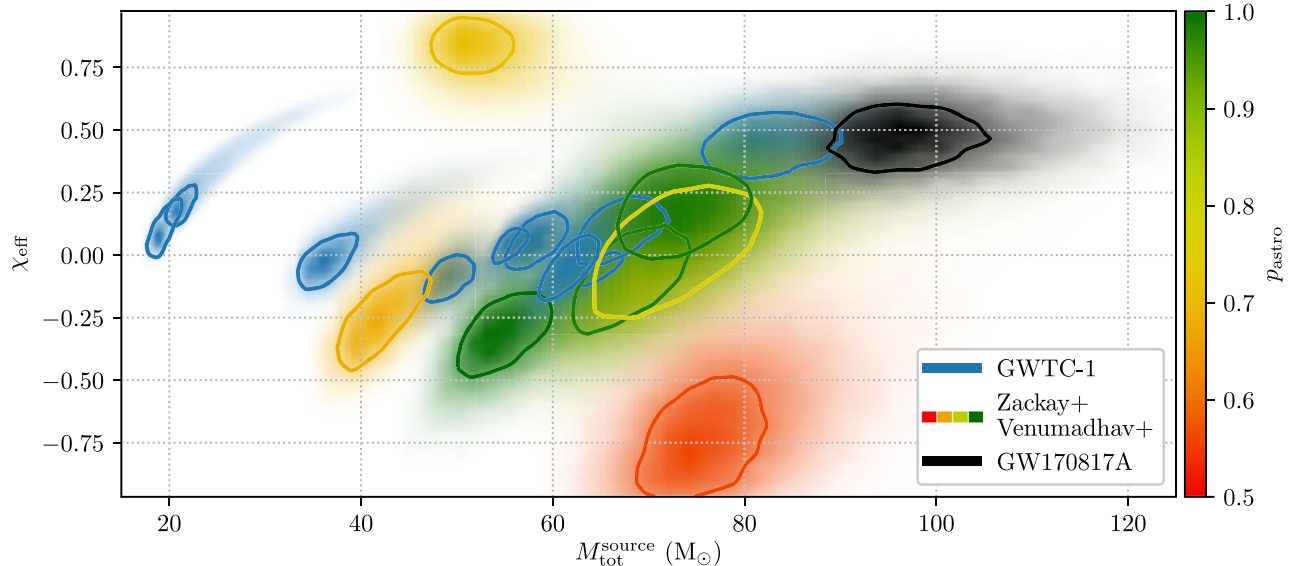
FIG. 7.   Binary black holes events reported from O1 and O2, in the plane of source-frame total mass vs effective spin. In blue are shown the 10 BBH events reported in GWTC-1 [1], all of them are certainly astrophysical in origin ($p_{astro} = 1$). Color coded by p_{astro} are shown 7 additional events with $p_{astro} > 0.5$ that our previous searches found [2,4]. In black we show GW170817A. Displayed are 1σ probability contours, i.e., enclosing $1 - e^{-1/2} \approx 0.39$ of the probability distribution.

inferred cutoff mass is at ∼42–44 $M_\odot$ (effectively the lowest end of the posterior of the most massive system, GW170729), with a tail toward higher values. If GW170817A were incorporated into such an analysis, the inferred cutoff would be at higher values of the mass and the constraint would be strengthened compared to using just GW170729 alone.

Finally, Fig. 7 places the event in the context of all the other events detected thus far in terms of the total source-frame mass and effective spin parameter, χ_{eff}. Interestingly, GW170817A follows the emerging trend of the total source-frame mass and the effective spin being correlated at the heavier end of the detected population of events. We need a proper accounting of the selection effects of the search pipelines to assess the astrophysical relevance of any putative correlations.

V. CONCLUSIONS

In this paper, we presented and applied a new method to assess the false-alarm rates of loud single detector triggers that have weak counterpart signals in other detectors. This method is motivated by the fact that there is significant sensitive volume in this regime when the detectors that make up the network have disparate sensitivities, such as was the case during the O2 run, and that analyses of coincident triggers do not cover this volume due to cuts on SNR. We also note that in the ongoing O3 run, the Livingston detector continues to be substantially more sensitive than the others, and hence we expect that the method we present will be

important even in the future. It could be especially interesting to apply this technique to several binary neutron star candidates identified thus far by the LVC during O3 that have only Livingston–Virgo codetection.[6]

When we apply our method to data from the O2 observing run, we detect two additional significant events. One of the events, GW170817A is the merger of a pair of very massive black holes; its estimated parameters suggest that it could be the most massive merger reported so far. It has been theoretically suggested that stellar mass BHs are subject to a mass cutoff at ∼40–50 $M_\odot$ due to the physics of pulsational pair instability supernovae and pair instability supernovae of the progenitor star [27,28]. GW170817A should be very valuable in constraining the existence and the exact location of such a mass cutoff [29–33].

The other event, GWC170402, if genuine, is perhaps the most interesting as it shows hints of additional physics not included in the waveform models we have used in this paper. We will present a detailed analysis of this event in a companion paper.

It is also interesting that the distribution of masses and spins of the events detected so far indicates a correlation between the total source-frame mass and the effective spin parameter, χ_{eff}. We need more detections and a careful population analysis to confirm the astrophysical nature of this correlation; if real, it may shed light on the binary stellar evolution of massive stars.

---

[6]https://gracedb.ligo.org/superevents/public/O3/.

APPENDIX A: INHOMOGENEOUS DISTRIBUTION OF GLITCHES

In Sec. II A, we provided a ranking of single-detector (L1) triggers. The ranking relied on the empirical observation that after we applied signal-quality vetoes to the triggers that the matched filtering procedure returned, the remaining glitches were confined to certain glitch-prone regions within the set of templates that we used.

The table and associated figure in Fig. 8 present evidence of this effect. The table in the left-hand panel shows the numbers of veto-passing triggers above three threshold values of SNR^2 (45, 55, and 65) for the heavier banks and sub-banks that we used in our search. Note that the

templates in bank BBH 2 and its subbanks cover signals with chirp-masses $m_c \in (12, 20)M_\odot$, while the search in this paper covers signals with $m_c > 20 \ M_\odot$. We include this bank because its subbank BBH (2,2) shows the most dramatic example of the phenomenon of localized glitches.

At low values of the SNR (the first column in the table), the numbers are controlled by Gaussian noise, and hence the disparity in numbers largely reflects the different numbers of templates in the various banks/subbanks (except BBH (4, 3) and (4, 4), which show signs of glitches even at SNR$^2 = 45$). At larger values of the SNR, the distributions are dominated by glitches, and we can see that the effects are localized to within a few subbanks. Even within subbanks, there are a few glitch-prone templates that dominate the tail of the distribution. The figure in the right-hand panel of Fig. 8 is a scatter plot of the first two coefficients that index our template bank for the triggers in BBH (2, 2). We see that almost all the glitches are localized to a small region within the bank (as shown by the red and black markers, which are the triggers with $\rho_L^2 > 55$ and 65, respectively. We further note that the bank with the highest rate of discovered BBH events, BBH (3,0), is essentially clean. This is evidenced by the dramatic drop between the number of triggers with $\rho_L^2 > 45$ and $\rho_L^2 > 55$ (a factor of 30–100, removing the number of detected (and expected) foreground, and accounting for Poisson errors). This factor is consistent with the one observed in bank BBH (2,0) which we infer to be clean as well (the lack of many detections above $\rho_L^2 > 65$ is a confirmation).

APPENDIX B: A SPURIOUS CANDIDATE

The ranking in Sec. II A marked the L1 triggers for all previous loud events, and produced a short list of remaining single-detector candidates. In Table I, we noted that the

Bank ID	# $\rho_L^2 > 45$	# $\rho_L^2 > 55$	# $\rho_L^2 > 65$
(2,0)	10178	172	1
(2,1)	1558	50	7
(2,2)	734	226	102
(3,0)	337	18	7[a]
(3,1)	157	11	3
(3,2)	41	8	4
(4,0)	37	3	1[b]
(4,1)	14	1	0
(4,2)	9	3	2
(4,3)	32	11	4
(4,4)	215	77	31

[a] Six of the seven triggers in bank (3,0) are previously declared gravitational wave signals. The seventh is declared in this paper
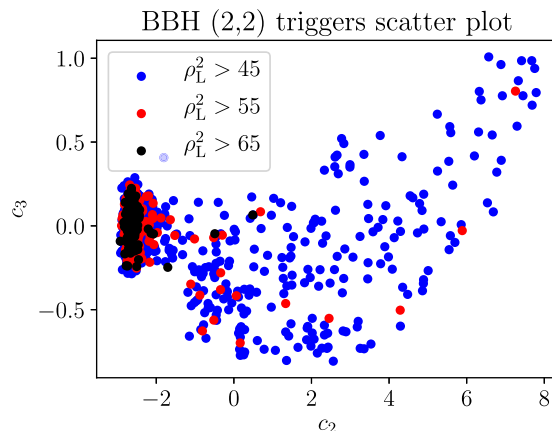[b] This trigger is GW170823

BBH (2,2) triggers scatter plot

FIG. 8. The table in the left-hand panel shows the number of veto-passing L1 triggers in each sub-bank above a few threshold values of the SNR. The non-uniform numbers of triggers with $\rho_L^2 > 65$ shows that glitches are localized within certain sub-banks. The plot in the right-hand panel shows the coefficients labeling the templates for triggers above the thresholds for bank BBH (2, 2). Note that glitches are localized within a small region of parameter space.
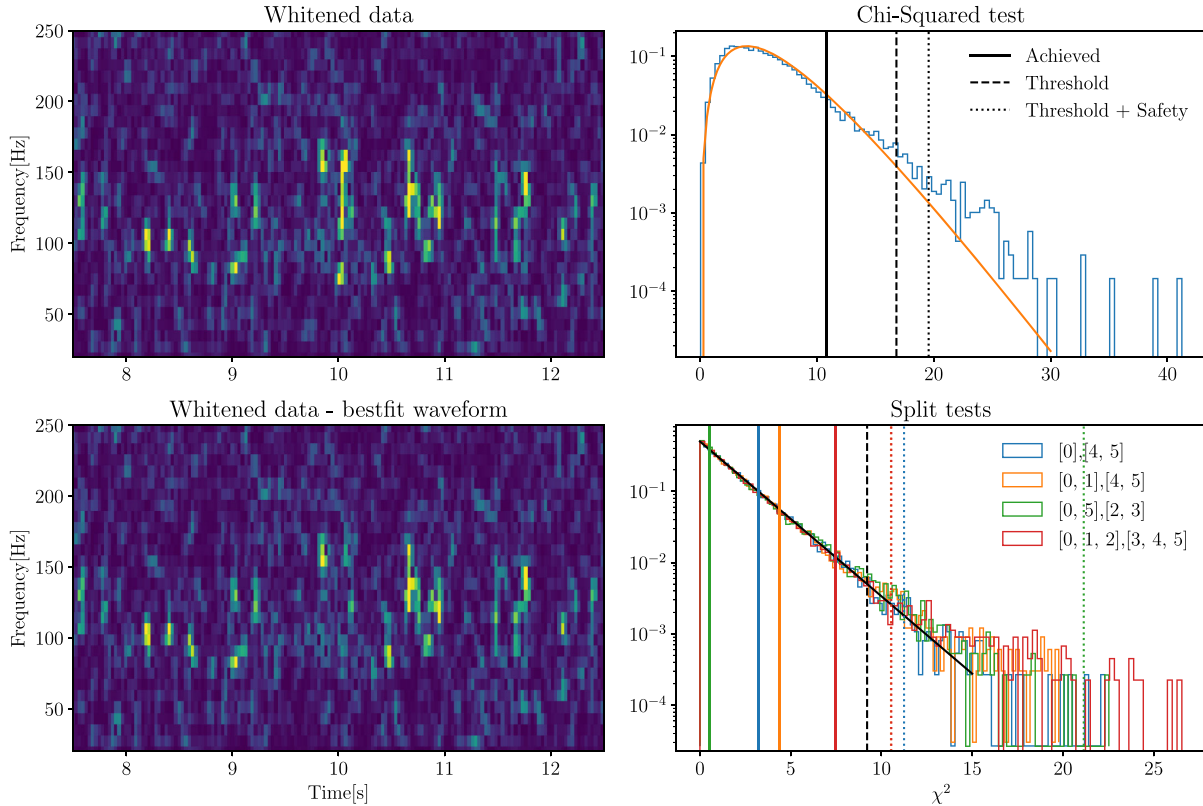
FIG. 9. Details of the glitch appearing in the last row of Table I. Top left: spectrogram of the data around the trigger, the merger time is at 10 seconds. It is visually obvious that the event is a glitch, due to the activity in the few seconds around it. Bottom left: spectrogram of the data with the best fit template removed. The chirp itself is of sufficiently high quality that the subtracted region cannot be rejected by itself, without the context of the surrounding data. Top right: results of a chi-squared veto based on the matched-filter overlaps of subsets of the whitened waveform. Bottom right: results of split-tests based on the matched-filter overlaps of a few combinations of subsets of the whitened waveform. Each color corresponds to a different test. See text for details. The histograms show off-event distributions of all the test-statistics we use to veto the trigger, and the solid lines show their predicted distributions. The solid vertical lines are the values of the test statistic that are achieved by the event, and the dotted vertical lines show the thresholds applied on the test statistics (triggers to the left are retained to avoid rejecting real events).

ranking procedure produced a candidate that had clear artifacts in its spectrogram. Figure 9 presents the spectrogram for the L1 data around the time of this candidate. Our automated pipeline relies on a series of tests to reject glitches, and Fig. 9 includes the results of these tests. The lower-left panel shows the test performed to check whether the signal-subtracted data shows excess power—since the nonstationarity persists on longer timescales, and the test checks against a local average, this candidate was not rejected. The right-hand panels show the results of the vetoes that test consistency between the matched-filtering scores of subchunks of the best-fit whitened waveform; the results for this candidate are within the thresholds that we impose based on our requirements not to veto real signals. We divide the best-fit whitened waveform into six chunks with equal values of SNR, and test for the consistency of the matched-filtering overlaps of these chunks. The top-right panel shows the results of a chi-squared-like test that tests consistency between all six chunks [35], and the

bottom-right panel shows the results of split tests that test consistency between certain combinations of the chunks (the designation $[a, \ldots], [b, \ldots]$ denotes tests in which we compare the set of overlaps $z_a, \ldots$ to the set of $z_b, \ldots$). More details of this procedure will be provided in a future paper [36].

While these tests help reduce the effects of glitches, they are not perfect since they were informed by our previous experiences looking at small subsets of the data. In principle, we can design a test with criteria such that we can better reject this candidate and any other candidates like it, and add it to the battery of tests we have. We choose not to do so, because we do not have several examples of this glitch to measure the selectivity of any tests, and more importantly, it would make our analysis less blind. In this particular case, even if the tests do not reject the candidate, there is enough obvious nonstationary behavior that we can visually reject the possibility that it is of astrophysical origin.

[1] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. X **9**, 031040 (2019).

[2] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, Phys. Rev. D **101**, 083030 (2020).

[3] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, Phys. Rev. D **100**, 023011 (2019).

[4] B. Zackay, T. Venumadhav, L. Dai, J. Roulet, and M. Zaldarriaga, Phys. Rev. D **100**, 023007 (2019).

[5] A. H. Nitz, T. Dent, G. S. Davies, S. Kumar *et al.*, arXiv: 1910.05331.

[6] L. Blackburn, L. Cadonati, S. Caride, S. Caudill *et al.*, Classical Quant. Grav. **25**, 184004 (2008).

[7] B. P. Abbott, R. Abbott, T. Abbott, M. Abernathy *et al.*, Classical Quant. Grav. **33**, 134001 (2016).

[8] M. Cabero *et al.*, Classical Quant. Grav. **36**, 155010 (2019).

[9] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **119**, 161101 (2017).

[10] J. Roulet, L. Dai, T. Venumadhav, B. Zackay, and M. Zaldarriaga, Phys. Rev. D **99** (2019).

[11] T. Dal Canton, S. Bhagwat, S. V. Dhurandhar, and A. Lundgren, Classical Quant. Grav. **31**, 015016 (2014).

[12] S. Bose, S. Dhurandhar, A. Gupta, and A. Lundgren, Phys. Rev. D **94**, 122004 (2016).

[13] B. Zackay, T. Venumadhav, J. Roulet, L. Dai, and M. Zaldarriaga, arXiv:1908.05644 [Phys. Rev. D (to be published)].

[14] A. H. Nitz, T. Dent, G. S. Davies, S. Kumar, C. D. Capano, I. Harry, S. Mozzon, L. Nuttall, A. Lundgren, and M. Tápai, Astrophys. J. **891**, 123 (2020).

[15] B. Zackay, L. Dai, and T. Venumadhav, arXiv:1806.08792.

[16] 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).

[17] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, Phys. Rev. D **93**, 044007 (2016).

[18] S. E. Woosley, Astrophys. J. **836**, 244 (2017).

[19] M. Spera and M. Mapelli, Mon. Not. R. Astron. Soc. **470**, 4739 (2017).

[20] M. Limongi and A. Chieffi, Astrophys. J. Suppl. Ser. **237**, 13 (2018).

[21] C. L. Rodriguez, M. Zevin, P. Amaro-Seoane, S. Chatterjee, K. Kremer, F. A. Rasio, and C. S. Ye, Phys. Rev. D **100**, 043027 (2019).

[22] M. Fishbach and D. E. Holz, Astrophys. J. **851**, L25 (2017).

[23] C. Talbot and E. Thrane, Astrophys. J. **856**, 173 (2018).

[24] D. Wysocki, J. Lange, and R. O'Shaughnessy, Phys. Rev. D **100**, 043012 (2019).

[25] J. Roulet and M. Zaldarriaga, Mon. Not. R. Astron. Soc. **484**, 4216 (2019).

[26] The LIGO Scientific and Virgo Collaborations, Astrophys. J. Lett. **882**, L24 (2019).

[27] W. A. Fowler and F. Hoyle, Astrophys. J. Suppl. Ser. **9**, 201 (1964).

[28] Z. Barkat, G. Rakavy, and N. Sack, Phys. Rev. Lett. **18**, 379 (1967).

[29] K. Belczynski *et al.*, Astron. Astrophys. **594**, A97 (2016).

[30] S. E. Woosley, Astrophys. J. **836**, 244 (2017).

[31] M. Spera and M. Mapelli, Mon. Not. R. Astron. Soc. **470**, 4739 (2017).

[32] P. Marchant, M. Renzo, R. Farmer, K. M. W. Pappas, R. E. Taam, S. de Mink, and V. Kalogera, Astrophys. J. **882**, 36 (2019).

[33] S. Stevenson, M. Sampson, J. Powell, A. Vigna-Gómez, C. J. Neijssel, D. Szécsi, and I. Mandel, Astrophys. J. **882**, 121 (2019).

[34] https://www.gw-openscience.org.

[35] B. Allen, Phys. Rev. D **71**, 062001 (2005).

[36] T. Venumadhav *et al.* (to be published).