Impact of noise transients on gravitational-wave burst detection efficiency of the *BayesWave* pipeline with multidetector networks

Yi Shuen C. Lee[1,2,*] Margaret Millhouse[1,2,3,†] and Andrew Melatos[1,2,‡]

[1]*School of Physics, The University of Melbourne, Victoria 3010, Australia*
[2]*Australian Research Council Centre of Excellence for Gravitational Wave Discovery (OzGrav),*
University of Melbourne, Parkville, Victoria 3010, Australia
[3]*Center for Relativistic Astrophysics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA*

Detection confidence of the source-agnostic gravitational-wave burst search pipeline *BayesWave* is quantified by the log signal-versus-glitch Bayes factor, $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$. A recent study shows that $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ increases with the number of detectors. However, the increasing frequency of non-Gaussian noise transients (glitches) in expanded detector networks is not accounted for in the study. Glitches can mimic or mask burst signals resulting in false alarm detections, consequently reducing detection confidence. This paper presents an empirical study on the impact of false alarms on the overall performance of *BayesWave*, with expanded detector networks. The noise background of *BayesWave* for the Hanford-Livingston (HL, two-detector) and Hanford-Livingston-Virgo (HLV, three-detector) networks are measured using a set of nonastrophysical background triggers from the first half of Advanced LIGO and Advanced Virgo's Third Observing Run (O3a). Efficiency curves are constructed by combining $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ of simulated binary black hole signals with the background measurements, to characterize *BayesWaves*'s detection efficiency as a function of the per-trigger false alarm probability. The HL and HLV network efficiency curves are shown to be similar. A separate analysis finds that detection significance of O3 gravitational-wave candidates as measured by *BayesWave* are also comparable for the HL and HLV networks. Consistent results from the two independent analyses suggests that the overall burst detection performance of *BayesWave* does not improve with the addition of Virgo at O3a sensitivity, because the increased false alarm probability offsets the advantage of higher $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$.

I. INTRODUCTION

The Advanced Laser Interferometer Gravitational-Wave Observatory (LIGO) [1] detectors in Hanford, Washington and Livingston, Louisiana, USA have completed three observing runs O1, O2 and O3 between 2015 and 2020, two of which were joint observations with the Advanced Virgo detector in Cascina, Italy [2]. The Kamioka Gravitational Wave Detector (KAGRA) [3–5] located in Hida, Japan also came online toward the end of O3, conducting a joint observation (O3GK) [6] with the GEO600 [7] detector in Hannover, Germany. As of the three observing runs, around 90 candidate gravitational wave (GW) events were collectively observed and reported in the Gravitational-wave Transient Catalogs (GWTCs) [8–11]. In May 2023, the LIGO-Virgo-KAGRA (LVK) collaboration began the fourth observing run O4 with the two LIGO detectors. The Virgo and KAGRA detectors are also expected to join O4 at a later date.

GW events observed so far by the LVK detectors are compact binary coalescences (CBCs), namely the mergers of binary black holes (BBH), binary neutron stars and neutron star-black hole binaries. CBCs are transient GW events, otherwise known as GW burst sources. Aside from CBCs, we expect to observe GW bursts from other astrophysical sources including but not limited to core-collapse supernovae [12,13], pulsar glitches [14], magnetar bursts[1] [15,16], nonlinear gravitational memory due to low-mass BBH mergers [17] and cosmic string cusps or kinks [18–20]. In addition, the possibility exists of GW bursts from astrophysical objects or processes that have not yet been discovered through electromagnetic observations. By their nature, GW waveforms of such novel signals are unclassified at present.

Traditionally, GW transient search pipelines use a matched filter [21–24] to compare the data to a bank of waveform templates obtained through various waveform modeling techniques [25–28]. Unlike CBCs, the

[*]ylee9@student.unimelb.edu.au
[†]meg.millhouse@gatech.edu
[‡]amelatos@unimelb.edu.au

---

[1]Magnetar bursts are short bursts ($\sim 0.1s$) of soft gamma-rays emitted by highly magnetized, isolated neutron stars. Their physical mechanism is unknown.

082002-1

waveforms of most prospective GW burst sources vary unpredictably from one event to the next and involve complicated physics beyond general relativity (e.g. hydrodynamics and neutrino transport). It is challenging to construct robust models with a few well-defined parameters which predict the waveforms, so template-based matched-filter searches for unmodeled GW bursts are impractical.

Several developed and emerging pipelines exist to perform source-agnostic GW burst searches [29–31], including but not limited to coherent WaveBurst (cWB) and *BayesWave* (BW). The cWB [32–35] burst search pipeline is used for offline analysis and online, low-latency generation of triggers for electromagnetic followups. Detection statistics of the cWB algorithm scales with the excess power in the time-frequency domain. BW uses the transdimensional reversible jump Markov chain Monte Carlo (RJMCMC) algorithm which adjusts the model dimension in response to the data. For that reason, BW is computationally intensive and is only used to follow-up potential GW candidates identified by other search pipelines. In the all-sky GW burst searches of the three Advanced LIGO and Advanced Virgo observing runs [36–38], cWB is used to analyse the full dataset and BW is used to follow-up cWB triggers [39–43]. Previous studies have shown that hierarchical implementation of cWB and BW enhances detection confidence [44].

As of O4, the LVK global network comprises of four large-scale detectors. With the commissioning of LIGO-India well under way [45], the network of GW detectors is expected to expand in the coming years. The expanding network of detectors with improved sensitivities increases the duty cycle, sky coverage and the accuracy of sky localization [46,47]. However, having more detectors also increases the susceptibility of the network to transient nonastrophysical disturbances, as noted in O3 [9,11,38,48]. These non-Gaussian instrumental noise transients, otherwise known as "glitches", appear as excess power in detector data and can mimic or mask unmodeled GW bursts. To enable high confidence detections with high astrophysical significance, glitches have to be identified and mitigated appropriately. Several efforts have been made to identify and characterize glitches by their origin and/or morphology [49–53]. Three common glitches in the LIGO-Virgo detectors are termed blip [54], whistle [55] and scattered light [56]. The whistle and scattered light glitches are of relatively longer duration ($\sim$0.7–2.0 s) and their origins are well understood. Blip glitches, on the other hand, are transient power spikes which lasts for $\sim$0.1 s and spans a wide frequency band ($\sim$$10^2$ Hz), typically of unknown origin. In cases where the glitch origin is unknown, further investigations are necessary before flagging a glitch and regressing it from the data to avoid overlooking astrophysical signals [48,57–64].

The BW algorithm enables the joint detection and characterization of GW burst and instrumental glitches, with no *a priori* assumptions of the source or morphology. Studies have been conducted to evaluate various aspects of BW's performance with multidetector networks, including detection confidence, parameter estimation and waveform reconstruction [44,65–68]. In Ref. [66], the detection confidence of BW with multidetector networks is quantified using the algorithm's detection statistic: the log signal-to-glitch Bayes factor, $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$. The study showed analytically that increasing the number of detectors in a network has a positive impact on $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$, following derivations in Ref. [65]. The results are verified empirically with simulated BBH signals. While the outcome is promising, the study does not consider the increase in glitch rate in an expanded detector network, i.e. it only focuses on the $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ of astrophysical events injected into simulated data in the absence of glitches. This paper generalizes Ref. [66], presenting a fuller analysis of BW's burst detection performance with expanded detector networks by accounting for the detector noise background using real detector data. For noise background measurements, we combine data from the first half of O3 (O3a) for the LIGO Hanford (H), LIGO Livingston (L) and Virgo (V) detectors, in particular the HL (two-detector) and HLV (three-detector) networks. We compare the overall performance of BW between the HL and HLV networks in O3a, noting that Virgo is less sensitive than HL; in contrast, the sensitivities of all three detectors may be comparable in future observing runs. The performance of BW is evaluated by comparing the $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ produced by astrophysical signals against the respective detector network backgrounds, using two independent injection sets. A set of simulated BBH signals is used to construct efficiency curves for characterizing BW's detection efficiency as a function of detection significance. To check for consistency, we analyse O3-like CBC signals to measure BW's detection significance of O3 GW candidates from GWTC-2 [9] and GWTC-3 [11].

The rest of this paper is organized as follows. In Sec. II we outline the key features of the BW algorithm. In Sec. III we discuss the datasets used to study BW's performance: (i) HL and HLV background triggers for background measurements, (ii) simulated BBH injections and (iii) O3-like CBC waveform injections. In Sec. IV we present BW's background measurements. In Sec. V we present the results for BW's efficiency analysis with the simulated BBH injections, and in Sec. VI the significance measurements for the O3 GW candidates. We summarize our findings and discuss avenues for future work in Sec. VII.

II. BAYESWAVE

In this section, we briefly overview the fundamental principles of the BW algorithm (Sec. II A), the models for

data reconstruction (Sec. II B) and the Bayes factor for model selection (Sec. II C).

A. Algorithm overview

The BW algorithm is designed to adaptively reconstruct of nonstationary and non-Gaussian transients in the data, using models with variable dimensions. The name of the algorithm, *BayesWave*, expresses two key concepts: (i) waveform reconstruction using sine-Gaussian (also known as Morlet-Gabor) wavelets, and (ii) the implementation of Bayesian inference to discriminate signals from glitches.

For a given detector i, the data $d_i(t)$ at time t consists of three components: the GW signal $h_i(t)$, which is bounded in t for burst sources; glitches $g_i(t)$, which are also bounded in t; and random detector noise $n_i(t)$, which is present continuously. That is, we have $d_i(t) = h_i(t) + g_i(t) + n_i(t)$. The BW algorithm attempts to reconstruct the transient, non-Gaussian features i.e., $h_i(t)$ and/or $g_i(t)$ in a stretch of detector data, by summing a set of sine-Gaussian wavelets. A single sine-Gaussian wavelet in the time domain takes the mathematical form

$$\Psi(t;\lambda) = Ae^{-(t-t_0)^2/\tau^2} \cos\left[2\pi f_0(t-t_0) + \phi_0\right], \quad (1)$$

with $\tau = Q/(2\pi f_0)$ and $\lambda = \{t_0, f_0, Q, A, \phi_0\}$. The symbols t_0, f_0, Q, A, ϕ_0 denote the central time, central frequency, quality factor, amplitude and phase offset of the wavelet respectively. Since the wavelets are not linearly independent, they form a frame and not a basis (see Sec. 3 of Ref. [41] for further details).

Wavelet parameters are sampled from designated prior distributions using the transdimensional reversible jump Markov chain Monte Carlo (RJMCMC) technique [69]. The implementation of transdimensional jumps allows for the number of wavelets to vary depending on the waveform complexity. By summing all the wavelets at each iteration of the RJMCMC chain, we obtain a posterior distribution of waveform models. For further details on wavelet parameter estimation and the measures taken to optimize convergence to the target distribution, we refer the reader to Refs. [41] and [42].

B. Modeling the data

BW reconstructs the detector data using three independent models, namely (i) the GW signal plus Gaussian-noise model, $\mathcal{S}$, (ii) glitches plus Gaussian-noise model, $\mathcal{G}$ (iii) Gaussian-noise model, $\mathcal{N}$. In this work, we are interested in the Bayes factor between the $\mathcal{S}$ and $\mathcal{G}$ models as a quantitative measure for BW's detection confidence.

1. Signal model, $\mathcal{S}$

Recall that the five intrinsic parameters of a sine-Gaussian wavelet can represented with a single parameter

vector $\lambda = \{t_0, f_0, Q, A, \phi_0\}$. If a real GW signal is present in the data of a multidetector network, we expect it to be coherent across all detectors in the network, albeit with different signal-to-noise ratio (SNR) and polarization per detector depending on the sensitivity and orientation of the detectors respectively. Therefore when reconstructing the data using the signal model, the same wavelet parameters are used across all detectors in the network. The set of intrinsic parameters for the signal model ($\mathcal{S}$) is given by $\lambda^{\mathcal{S}} = \lambda_1 \cup \lambda_2 \ldots \cup \lambda_{N^{\mathcal{S}}}$, where $N^{\mathcal{S}}$ denotes the number of wavelets used in the signal reconstruction. These parameters are geocentric, meaning they are measured at a reference point located at the center of the Earth.

Since the signal models represent astrophysical GW signals, all $N^{\mathcal{S}}$ wavelets used in the reconstruction also share a set of extrinsic parameters $\mathbf{\Omega} = \{\theta, \phi, \epsilon, \psi\}$. The symbols denote the right ascension, declination, ellipticity and polarization angle of the GW in order of appearance. The complete set of signal model parameters is then given by $\boldsymbol{\theta}^{\mathcal{S}} = \lambda^{\mathcal{S}} \cup \mathbf{\Omega}$.

The geocentrically measured signal waveforms, parameterised by $\lambda^{\mathcal{S}}$, can be projected onto the ith detector using the detector's unique time delay operator $\Delta t_i(\theta, \phi)$, along with the antenna beam pattern response functions $F_i^+(\theta, \phi, \psi)$ and $F_i^+(\theta, \phi, \psi)$ of the plus (+) and cross (×) polarizations.[2] Mathematically we write [65,70]

$$h_i(f; \lambda^{\mathcal{S}}, \mathbf{\Omega}, N^{\mathcal{S}}) = (F_i^+ \tilde{h}_+ + F_i^\times \tilde{h}_\times)e^{2\pi i f \Delta t_i}, \quad (2)$$

where $\tilde{h}_p$ denotes the Fourier transform of the time domain geocentric GW signal, $h_p(t)$ for polarization p. The version of BW used in our analysis assumes elliptical polarization such that the ellipticity parameter ϵ maps $\tilde{h}_+$ to the cross polarization $\tilde{h}_\times$ via

$$\tilde{h}_\times = \epsilon \tilde{h}_+ e^{i\pi/2}, \quad (3)$$

and $\tilde{h}_+$ is expressed as a linear combination of sine-Gaussian wavelets in the frequency domain (obtained by taking the Fourier transform of Eq. (1):

$$\tilde{h}_+(f) = \sum_{n=1}^{N^{\mathcal{S}}} \tilde{\Psi}(f; \lambda_n). \quad (4)$$

2. Glitch model, $\mathcal{G}$

Unlike GW signals, instrumental glitches and noise are uncorrelated across the detector network. Therefore the glitch model uses independent sets of wavelets to

---

[2]Antenna pattern functions are typically a function of time. However, the time dependence is omitted here with the assumption that the antenna patterns are constant over the short duration of GW burst. This assumption is conventional across all burst searches.

reconstruct glitches in each detector. Let $N^{\mathcal{G}_i}$ denote the number of wavelets and $\lambda^{\mathcal{G}_i} = \lambda_1^i \cup \lambda_2^i \ldots \cup \lambda_{N^{\mathcal{G}_i}}^i$ be the set of wavelet parameters used in the glitch model reconstruction of detector i. We can then write the glitch model for the ith detector as

$$g(\lambda^{\mathcal{G}_i}, N^{\mathcal{G}_i}) = \sum_{n=1}^{N^{\mathcal{G}_i}} \tilde{\Psi}(f; \lambda_n^i). \qquad (5)$$

Thus for a network with $\mathcal{I}$ detectors, the complete set of glitch model wavelet parameters is given by $\boldsymbol{\theta}^{\mathcal{G}} = \lambda^{\mathcal{G}_1} \cup \lambda^{\mathcal{G}_2} \ldots \cup \lambda^{\mathcal{G}_{\mathcal{I}}}$. Note that there are no extrinsic parameters in the glitch model as it assumes the non-Gaussianity in the data to be independent in each detector (i.e. nonastrophysical) [65].

3. Gaussian-noise model, $\mathcal{N}$

In contrast to $\mathcal{S}$ and $\mathcal{G}$ which models non-Gaussian transient components of the detector data, the *BayesLine* algorithm [40] is implemented within BW to model the Gaussian-noise power spectral density (PSD). LIGO and Virgo Gaussian noise sources can be classified into three broad frequency bands: (i) seismic noise ($\sim$10 Hz), (ii) thermal noise ($\sim$10–200 Hz) and (iii) quantum (photon) shot noise ($\gtrsim$200 Hz). Moreover, various aspects of the detector apparatus including mirror suspensions, calibration lines, and the AC electrical supply are recurrent sources of high-power, narrow-band spectral lines. *BayesLine* collectively characterizes these noise features by modeling the PSD using cubic splines and Lorentzians as bases to fit smooth broad-band noise and narrow-band linelike features respectively. The mathematical details of *BayesLine* are incidental to this paper, a full description can be found in [40].

By amalgamating all plausible PSD models from *BayesLine*, we obtain $\mathcal{N}$. As its name suggests, $\mathcal{N}$ models the data as purely Gaussian noise. In fact $\mathcal{S}$ and $\mathcal{G}$ also incorporate *BayesLine* for PSD estimation on top of the wavelet models for non-Gaussian feature(s), and are therefore known as composite models [41].

C. Bayesian model selection

BW compares the model evidences via the Bayes factor to give the relative odds between the hypotheses described in Sec. II B. For a given model $\mathcal{M}$, the evidence is calculated by

$$p(\boldsymbol{d}|\mathcal{M}) = \int d\boldsymbol{\theta}^{\mathcal{M}} p(\boldsymbol{\theta}^{\mathcal{M}}|\mathcal{M}) p(\boldsymbol{d}|\boldsymbol{\theta}^{\mathcal{M}}, \mathcal{M}) \qquad (6)$$

where $p(\boldsymbol{\theta}^{\mathcal{M}}|\mathcal{M})$ is the prior i.e the probability that $\mathcal{M}$ is parametrized by $\boldsymbol{\theta}^{\mathcal{M}}$ prior to observation of the data $\boldsymbol{d}$; and $p(\boldsymbol{d}|\boldsymbol{\theta}^{\mathcal{M}}, \mathcal{M})$ is the likelihood of observing $\boldsymbol{d}$ given $\boldsymbol{\theta}^{\mathcal{M}}$. In essence, the evidence is the likelihood of producing the data $\boldsymbol{d}$ from the hypothesis $\mathcal{M}$ marginalised over the parameter

space of $\boldsymbol{\theta}^{\mathcal{M}}$, thus it is otherwise known as the marginalised likelihood [65]. Obtaining model evidences directly from the integral in Eq. (6) is computationally expensive, especially for complex and highly parameterised models. Therefore BW combines the parallel-tempered RJMCMC algorithm [71] and thermodynamic integration [72] to compute the evidences. Implementations of these methods are detailed in Refs. [41] and [40].

The Bayes factor between two models, $\mathcal{M}_\alpha$ and $\mathcal{M}_\beta$, is the ratio of their evidences:

$$\mathcal{B}_{\alpha,\beta}(\boldsymbol{d}) = \frac{p(\boldsymbol{d}|\mathcal{M}_\alpha)}{p(\boldsymbol{d}|\mathcal{M}_\beta)}. \qquad (7)$$

$\mathcal{B}_{\alpha,\beta}(\boldsymbol{d}) > 1$ suggests that $\mathcal{M}_\alpha$ is more strongly supported by the data and vice versa. The Bayes factor inherently considers model complexity in model selection by penalizing over-fitting. This is a corollary of Occam's razor, which prefers simplicity over complexity among competing models. Occam's razor is not deliberately implemented; rather it is an inherent consequence of using Bayes' theorem and enters via the parameter space volume in Eq. (6). For a detailed mathematical interpretation, we refer the reader to Sec. IV A of [66].

A study conducted by Littenberg *et al.* [65] to assess BW's ability to distinguish between GW signals and instrumental glitches shows that for a two-detector network with interferometers of equal sensitivity (i.e., the HL network), the primary scaling of the Bayes factor goes as [65]

$$\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} \propto N \ln(\text{SNR}_{\text{net}}). \qquad (8)$$

A simplifying assumption is that the number of wavelets used in the signal model $\mathcal{S}$ is the same as the glitch model $\mathcal{G}$ for a single detector, viz. $N^{\mathcal{S}} = N^{\mathcal{G}_i} = N$. For a network with $\mathcal{I}$ detectors, the overall network SNR of the non-Gaussian transient in the data is given by

$$\text{SNR}_{\text{net}}^2 = \sum_{i=1}^{\mathcal{I}} \text{SNR}_i^2 \qquad (9)$$

where SNR_i is the SNR in detector i. Altogether, Eq. (8) suggests that $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ and hence detection confidence scale with both signal strength and waveform complexity.

In a complementary study [66], BW is used to recover injected BBH signals from the HL, HLV and HLKV networks to quantify its detection confidence with expanded detector networks. In this study, $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ is further shown to scale with the number of detectors in the network, $\mathcal{I}$, according to

$$\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} \propto \mathcal{I} N \ln(\text{SNR}_{\text{net}}). \qquad (10)$$

In other words, BW's detection confidence is directly and positively impacted by increasing the number of detectors in the network, all else being equal.

III. BAYESWAVE EFFICIENCY ANALYSIS

In standard GW searches, the astrophysical significance of a detection candidate is determined by the frequency of false alarms. False alarms are nonastrophysical events with detection statistics corresponding to that of GW candidates. To estimate the prevalence of false alarms, one can count the number of triggers produced by the detector background which does not contain astrophysical signals.

Ref. [66] assesses BW's detection confidence for expanded detector networks using only the detection statistic $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ produced by astrophysical events. However as the global detector network expands, the likelihood of instrumental glitches increases. The associated increase in false alarm detections reduces astrophysical significance of detections, thereby reducing detection confidence. Unmodeled burst searches (e.g., with BW) place fewer constraints on the waveform morphology, and are therefore confounded more readily by glitches compared to modeled searches (e.g., with a matched filter) [73–77]. Since the significance of $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ is influenced by false alarms, we present a more complete analysis of BW's performance with expanded detector networks by considering the impact of detector noise backgrounds on detection confidence.

We use detection efficiency P_{det} as a figure of merit to compare the overall performance of BW between the HL (two-detector) and HLV (three-detector) networks. P_{det} is typically characterised as a function of detection significance by means of a receiver-operating-characteristic (ROC) curve,[3] also known as an efficiency curve. In this study, we use the per-trigger false alarm probability P_{FA} as a measure of significance. We define P_{FA} as the probability that a trigger measured with a given detection statistic is a false alarm, and P_{det} as the probability of detecting an astrophysical event with a given significance. Higher P_{FA} indicates low astrophysical significance. Therefore higher P_{det} is achieved, if higher P_{FA} is tolerated. P_{FA} should not be confused with the false alarm rate (FAR), which measures the number of false alarms per unit time. We further discuss the difference in Sec. IV, and explain why we use P_{FA} instead of FAR in our analysis.

In order to measure P_{FA}, we need to understand the distribution of $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ produced by the detector noise background. It is challenging to construct models that can accurately predict the noise background, so we empirically obtain the background distribution by applying BW to triggers identified by cWB from time-shifted background data of the HL and HLV networks. The distribution of $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ produced by the background triggers is then used to compute P_{FA}. Using BW to recover a population of injected signals, we obtain a distribution of $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ for astrophysical events. Combining $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ of background triggers and

astrophysical injections, we compute P_{det} as a function of P_{FA} to construct efficiency curves. We discuss the methods of constructing efficiency curves in greater detail in Sec. VA.

To study the impact of the noise background on BW's overall performance with expanded detector networks, we compare the efficiency curves between the HL and HLV networks for a synthesised population of BBHs. As a consistency check, we also analyse a set of O3-like CBC waveforms to measure and compare BW's detection significance of O3 GW detection events for HL and HLV. In the following sections, we detail the background and injection datasets for the analyses.

A. Background data

In GW data analysis, it is standard practice to use the time-shifting method to create pseudo-real detector datasets for noise background estimations [36–38,78–80]. The time-shifting method introduces artificial time off-sets between the outputs of GW detectors operating in concert. The offsets are much larger than the coherence time (~ 10 ms) of any real GW signals between the detectors, determined by the distance between the detectors and the GW propagation speed. As a result, coincident triggers in the time-shifted data cannot be astrophysical. By performing time-shifts repetitively on months worth of detector data, we obtain an artificially extended set of background data with effective livetimes[4] spanning thousands of years. We can use this to estimate P_{FA} by empirically measuring the fraction of background (i.e. noise-induced) triggers above a selected detection threshold in the time-shifted background [78].

In the Advanced LIGO and Advanced Virgo all-sky searches for short GW bursts [36–38], the cWB algorithm is used to analyse the full observational data. Due to the implementation of RJMCMC, BW is computationally intensive. Thus, BW is only used to follow-up subsets of cWB triggers. Although Ref. [44] has shown that the hierarchical implementation of cWB and BW enhances detection confidence in all-sky burst searches, the aim of our study is to assess the independent burst detection performance of BW. By convention, we use preexisting trigger lists generated by the cWB pipeline [38,81] to downselect triggers for BW background measurements, but we do not make any claims on cWB's background and detection efficiency. We choose to use the trigger list for the first half of O3 (O3a), acquired from the cWB low-frequency (16–1024 Hz) all-sky analysis of the full time-shifted O3a background data. The analysis is conducted separately for the HL and HLV networks, on background data obtained by applying time-shifts on 104.94-day (HL) and 75.19-day (HLV) segments of the

---

[3]Typical ROC curves plot probability of detection (true positives) on the vertical axis and probability of false alarm (false positives) on the horizontal axis.

[4]The extended time interval obtained as a result of time-shifting is known as the effective livetime.

real-time O3a detector data. The time-shifted background data accumulates 981 years and 573 years of effective livetimes for the HL and HLV networks respectively.

We select triggers by thresholding their cWB detection statistic, ρ, which scales with the network SNR of the signal present in the data [32,33]. We arbitrarily nominate $\rho_{\text{threshold}} = 7$ as the significance threshold, in line with previous work [38,44]. Triggers below the threshold are presumed to have insignificant impacts on detection efficiency and therefore excluded from the noise background measurement. cWB identifies 2×10^3 and 7×10^3 triggers[5] with $\rho > \rho_{\text{threshold}}$ in the HL and HLV background datasets respectively, but thousands of triggers are still too expensive to handle computationally in this paper. A straightforward approach is to increase $\rho_{\text{threshold}}$, but that would deliberately exclude low-SNR triggers from the background measurement. To avoid implementing a stricter $\rho_{\text{threshold}}$, we run BW on a fraction (denoted by X) of randomly selected triggers from the full trigger list, all of which satisfy $\rho > \rho_{\text{threshold}} = 7$. We set $X = 0.45$ and $X = 0.15$ for the HL and HLV datasets respectively to deliver roughly equal numbers of triggers from the two networks. The reduced HL and HLV background datasets consist of 1008 and 1134 triggers respectively. We employ BW to analyse the datasets to obtain $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ for each background trigger. The BW analysis uses the same settings as the Advanced LIGO and Advanced Virgo O3 all-sky search for short GW bursts [38] (see Appendix A).

From the BW analysis, we flag background triggers that are more consistent with the pure Gaussian-noise model, $\mathcal{N}$, than the composite signal plus Gaussian-noise model $\mathcal{S}$. By definition, the $\ln \mathcal{B}_{\mathcal{S},\mathcal{N}}$ error bars of these triggers encompass values less than or equal to zero, i.e.,

$$\ln \mathcal{B}_{\mathcal{S},\mathcal{N}} - \Delta \ln \mathcal{B}_{\mathcal{S},\mathcal{N}} \leq 0 \qquad (11)$$

where $\Delta \ln \mathcal{B}_{\mathcal{S},\mathcal{N}} = \sqrt{[\Delta \ln p(\boldsymbol{d}|\mathcal{S})]^2 + [\Delta \ln p(\boldsymbol{d}|\mathcal{N})]^2}$ is the width of the error bars and $\Delta \ln p(\boldsymbol{d}|\mathcal{M})$ denotes the uncertainty of the log evidence of model $\mathcal{M}$ [41]. For these Gaussian-noiselike triggers, the $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ is meaningless as it serves to compare the evidences of models that characterize non-Gaussianity. Nevertheless, these triggers cannot be discarded from the background measurement as they satisfy $\rho > \rho_{\text{threshold}}$. Therefore we assign them with an arbitrarily low detection statistic, $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} = -500$, to indicate minimal astrophysical significance. A total of 268 (218) out of 1008 (1134) HL (HLV) triggers are assigned $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} = -500$.

We present and discuss the background measurements in Sec. IV.

---

[5]These counts include triggers from all three search bins used in the cWB O3a low-frequency burst analysis: LF1, LF2, and LF3. The bins are classified based on trigger morphologies. Classification details can be found in [38] and [82].

B. Injections

In addition to the background measurement, the detection statistic distribution for astrophysical signals is required to evaluate BW's burst detection performance. We inject waveforms of known morphology and recover them using BW to empirically measure the distribution of $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$. Since CBCs are well understood, we use them in this study to assess BW's independent performance with HL and HLV. We analyse two different source populations: injection set 1 (IS1), comprising phenomenological BBH waveforms with fixed component masses but uniformly distributed SNR and extrinsic parameters, and injection set 2 (IS2), comprising CBC waveforms with parameters that resemble real GW events from O3. The following two subsections describe the objective and properties of each injection dataset in order.

1. Phenomenological BBH waveforms (IS1)

IS1 consists of simulated BBH waveforms with a choice of parameter space encompassing the range detectable by the Advanced LIGO and Advanced Virgo detectors. The waveforms are added to temporally spread out segments of HL and HLV data across all of O3a to reflect practical observation intervals. We use IS1 to characterize BW's detection efficiency (P_{det}) as a function of detection significance (P_{FA}) via efficiency curves and compare the performance of BW with the HL and HLV networks.

IS1 copies the injection set described in Sec. V of Ref. [66]. It consists of 1200 simulated BBH waveforms phenomenologically modeled using the IMRPhenomD [83,84] approximant. The BBH sources are nonspinning, nonprecessing and have equal component masses of $30 M_\odot$. They also have uniformly distributed sky locations, inclinations and polarization angles. The distances are randomly sampled such that the signal amplitude is detectable in simulated HLV data with network signal-to-noise ratio within range $10 \leq \text{SNR}_{\text{net}} \leq 50$. We use the same injection dataset for both the HL and HLV networks; we simply exclude Virgo data in the HL analysis. By Eq. (9), we expect SNR_{net} of any given event to be lower in the HL network compared to HLV.

The analysis in Ref. [66] injects and recovers waveforms using projected (simulated) O4 detector data. However, BW's background measurements for HL and HLV in this study are carried out using O3a background triggers as discussed in Sec. III A. In order to measure the P_{det} as a function of P_{FA}, detection statistics ($\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$) of the astrophysical signals must be compared with the background triggers of the same detector data. Thus we inject IS1 into arbitrarily selected segments of HL and HLV data throughout O3a. The O3a strain data is publicly available at the Gravitational Wave Open Science Centre (GWOSC) [85,86] and Fig. 2 of Ref. [9] shows representative amplitude spectral densities of the detectors. As with the
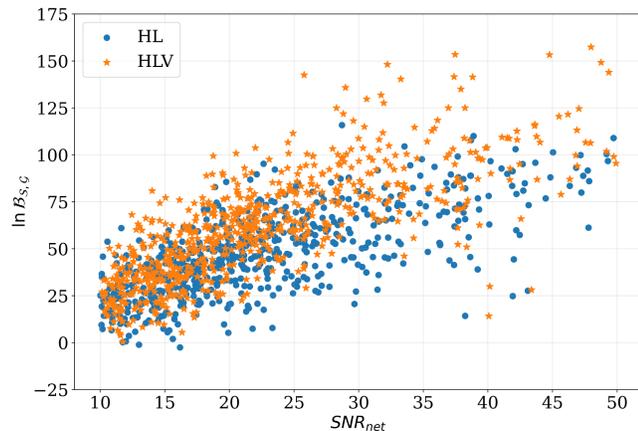
FIG. 1. Log signal-to-glitch Bayes factor $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ versus network signal-to-noise ratio SNR_{net} for IS1. The blue circles (orange stars) correspond to HL (HLV) network injections; each data point corresponds to a single injection. Gaussian-noise-like events with $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} = -500$ are not shown.

background, IS1 is analyzed using the same BW settings as Ref. [38].

Events of IS1 are injected into O3a data with the same distances sampled from the simulated data. Since O3a data is noisier and has a different characteristic PSD compared to the simulated HLV data, the SNR_{net} of IS1 events when injected into O3a data is lower than the referenced range $10 \leq SNR_{net} \leq 50$. In order to assess BW's performance under conditions relevant to practical searches, events below a designated detection threshold must be eliminated from the injection data. This is because they cannot serve as triggers by definition, in the same way that background triggers with $\rho < \rho_{threshold} = 7$ do not count as false alarms. Since this a designated search to assess the stand-alone efficiency of the BW algorithm, independent of cWB, we set a nominal significance threshold of $SNR_{cut-off} = 10$ for BW viz. only injection events with $SNR_{net} \geq SNR_{cut-off}$ in *both* the HL and HLV networks are adequately significant to be followed-up by BW and included in the efficiency curve analysis. Out of 1200 injections, 412 nondetection events are filtered out from IS1, leaving 788 events going forward.

From the remaining 788 events, BW identifies 157 (89) events consistent with Gaussian noise in the HL (HLV) network according to the $\ln \mathcal{B}_{\mathcal{S},\mathcal{N}}$ constraint defined in Section III A. These events are retained in the analysis dataset since they satisfy $SNR_{net} \geq 10$ but as with the background triggers, they are assigned $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} = -500$ to indicate low detection significance.

To show the overall distribution of IS1 events, we plot $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ versus SNR_{net} for the HL (blue circles) and HLV (orange stars) injections in Fig. 1. The plot shows all but the $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} = -500$ events to focus on events with astrophysically relevant $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$. Injections with comparable SNR_{net} are evidently recovered with higher $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ in HLV

compared to HL. This observation is consistent with Ref. [66] where $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ is analytically and empirically shown to increase primarily with $\mathcal{I}$.

Despite the astrophysical origin of IS1, *BayesWave* recovers two of the HL events with $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} < 0 \neq -500$ in Fig. 1, suggesting that the evidence for the 'incoherent' glitch model ($\mathcal{G}$) is higher than for the "coherent" signal model ($\mathcal{S}$). These events are also not consistent with Gaussian-noise, i.e., they have $\ln \mathcal{B}_{\mathcal{S},\mathcal{N}} > 0$. This is because the injected signal power in the frequency domain is only marginally above the sensitivity threshold in one detector, and is approximately one order of magnitude lower in the other. Equation (2) shows that the sensitivity of each detector to different sky locations, at a given time, depends on the antenna pattern functions. Therefore the $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} < 0$ recovery of the two HL injections, caused by the signal power imbalance across the detectors, is an inadvertent result of mismatched detector sensitivities to the randomly sampled sky locations at the time of injection. With additional coherent signal power from Virgo, the HLV-equivalents of these two events are recovered with $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} \sim 10^{1}$. This argument also applies to IS2 injections, discussed in Sec. III B 2.

We present the results of BW detection efficiency analysis with IS1 in Sec. V.

2. O3-like CBC waveforms (IS2)

To check for consistency with IS1, we measure BW's detection significance for real GW detection events in terms of P_{FA}, and compare the measurements between the HL and HLV networks. For this purpose, we implement BW on IS2 consisting CBC waveforms resembling O3a and O3b GW events from GWTC-2 [9] and GWTC-3 [11] respectively, otherwise known as off-source injected waveforms.

In IS1 the BBH waveforms are sampled from a fixed parameter space and added to detector data spread out across all of O3a; in IS2 the off-source injections are sampled from the matched-filter source parameter posteriors for GW detection events and added into the background data around the event epoch. Off-source injections are used in the GWTCs to test the consistency[6] between matched-filter (template-based) CBC waveforms and minimally modeled waveform reconstructions (e.g., cWB and BW) [8,9,11].

---

[6]Consistency test are performed by comparing the on-source and off-source match. On-source waveforms are reconstructed directly from the event data. The match, defined by $\mathcal{O} = \langle h_1 | h_2 \rangle / \sqrt{\langle h_1 | h_1 \rangle \langle h_2 | h_2 \rangle}$, measures the overlap between two waveforms h_1 and h_2. $\langle \cdot | \cdot \rangle$ is the noise-weighted inner product [87]. On-source match compares the maximum likelihood waveform from template-based parameter estimation of the actual event with the point estimate from minimally modeled reconstructions; off-source match compares the off-source injections with their respective reconstructions.

TABLE I. List of O3 GW events used to generate the off-source injections of IS2. The columns from left to right show: (i) The LIGO-Virgo-KAGRA (LVK) observing run in which the event was detected, (ii) event name, (iii) primary component mass m_1, (iv) secondary component mass m_2, (v) HLV network matched-filter SNR[†] and (vi) number of off-source injections (out of 50) that satisfy $\mathrm{SNR_{net}} \geq \mathrm{SNR_{cut\text{-}off}}$ and retained in IS2. Source parameter values displayed in the table are the median and the 90% symmertric credible intervals of the Bayesian posterior. Information in this table is copied directly from Table VI of GWTC-2 [9] (O3a events) and Table IV of GWTC-3 [11] (O3b events). [†]The network matched-filter SNR in this table is not to be confused with $\mathrm{SNR_{net}}$ which denotes *injected* network SNR of IS1 and IS2 events.

LVK run	Event name	m_1 ($M_\odot$)	m_2 ($M_\odot$)	Network SNR[†]	Number of off-source injections in IS2
O3a	GW190408_181802	$24.6^{+5.1}_{-3.4}$	$18.4^{+3.3}_{-3.6}$	$15.3^{+0.2}_{-0.3}$	50
O3a	GW190412	$30.1^{+4.7}_{-5.1}$	$8.3^{+1.6}_{-0.9}$	$18.9^{+0.2}_{-0.3}$	48
O3a	GW190503_185404	$43.3^{+9.2}_{-8.1}$	$28.4^{+7.7}_{-8.0}$	$12.4^{+0.2}_{-0.3}$	47
O3a	GW190512_180714	$23.3^{+5.3}_{-5.8}$	$12.6^{+3.6}_{-2.5}$	$12.2^{+0.2}_{-0.4}$	40
O3a	GW190513_205428	$35.7^{+9.5}_{-9.2}$	$18.0^{+7.7}_{-4.1}$	$12.9^{+0.3}_{-0.4}$	49
O3a	GW190517_055101	$37.4^{+11.7}_{-7.6}$	$25.3^{+7.0}_{-7.3}$	$10.7^{+0.4}_{-0.6}$	21
O3a	GW190519_153544	$66.0^{+10.7}_{-12.0}$	$40.5^{+11.0}_{-11.1}$	$15.6^{+0.2}_{-0.3}$	48
O3a	GW190521	$95.3^{+28.7}_{-18.9}$	$69.0^{+22.7}_{-23.1}$	$14.2^{+0.3}_{-0.3}$	44
O3a	GW190602_175927	$69.1^{+15.7}_{-13.0}$	$47.8^{+14.3}_{-17.4}$	$12.8^{+0.2}_{-0.3}$	44
O3a	GW190706_222641	$67.0^{+14.6}_{-16.2}$	$38.2^{+14.6}_{-13.3}$	$12.6^{+0.2}_{-0.4}$	41
O3a	GW190720_000836	$13.4^{+6.7}_{-3.0}$	$7.8^{+2.3}_{-2.2}$	$11.0^{+0.3}_{-0.7}$	24
O3a	GW190727_060333	$38.0^{+9.5}_{-6.2}$	$29.4^{+7.1}_{-8.4}$	$11.9^{+0.3}_{-0.5}$	49
O3a	GW190728_064510	$12.3^{+7.2}_{-2.2}$	$8.1^{+1.7}_{-2.6}$	$13.0^{+0.2}_{-0.4}$	48
O3a	GW190828_063405	$32.1^{+5.8}_{-4.0}$	$26.2^{+4.6}_{-4.8}$	$16.2^{+0.2}_{-0.3}$	48
O3a	GW190828_065509	$24.1^{+7.0}_{-7.2}$	$10.2^{+3.6}_{-2.1}$	$10.0^{+0.3}_{-0.5}$	19
O3a	GW190915_235702	$35.3^{+9.5}_{-6.4}$	$24.4^{+5.6}_{-6.1}$	$13.6^{+0.2}_{-0.3}$	47
O3a	GW190924_021846	$8.9^{+7.0}_{-2.0}$	$5.0^{+1.4}_{-1.9}$	$11.5^{+0.3}_{-0.4}$	36
O3b	GW200129_065458	$34.5^{+9.9}_{-3.2}$	$28.9^{+3.4}_{-9.3}$	$26.8^{+0.2}_{-0.2}$	50
O3b	GW200202_154313	$10.1^{+3.5}_{-1.4}$	$7.3^{+1.1}_{-1.7}$	$10.8^{+0.2}_{-0.4}$	35
O3b	GW200219_094415	$37.5^{+10.1}_{-6.9}$	$27.9^{+7.4}_{-8.4}$	$10.7^{+0.3}_{-0.5}$	13
O3b	GW200224_222234	$40.0^{+6.9}_{-4.5}$	$32.5^{+5.0}_{-7.2}$	$20.0^{+0.2}_{-0.2}$	36
O3b	GW200311_115853	$34.2^{+6.4}_{-3.8}$	$27.7^{+4.1}_{-5.9}$	$17.8^{+0.2}_{-0.2}$	48

IS2 comprises off-source injections of 22 independent GW events detected by the HLV network in O3. We summarize the relevant event properties in Table I. All events listed in Table I, except for GW200202_154313, are BW waveform consistency test candidates [11,11]. GW200202_154313 is excluded from the GWTC-3 consistency test due to low on-source match, but since the off-source injections for this event are available there is no reason to exclude it from IS2 for the assessment of BW's detection significance. A set of 200 off-source injections is available for each of the 22 GW events [9,11]. We arbitrarily select 50 out of the 200 off-source injections for each GW event, totaling $22 \times 50 = 1100$ injections in IS2. Even though a fraction of the GW events are O3b detections, we inject all off-source events into segments of O3a HLV data to ensure comparability with the O3a noise

background described in Sec. III A. The HL data are equivalent to the HLV data with Virgo removed.

As with IS1, only injections above the BW significance threshold are retained in IS2. The last column in Table I shows the number of off-source injections that exceeds the significance threshold, i.e., $\mathrm{SNR_{net}} \geq \mathrm{SNR_{cut\text{-}off}} = 10$ for each GW event. There are four GW events with less than 25 off-source injections (i.e., < 50%) satisfying the significance threshold, namely: GW190517_055101, GW190720_000836, GW190828_065509, and GW200219_094415. Since the $\mathrm{SNR_{net}}$ of off-source injections for these four events are sampled from match-filter network SNR posteriors with medians $\lesssim 11$ (see Table I), they are less likely to satisfy $\mathrm{SNR_{net}} \geq 10$. Assessments of astrophysical significance for GW events with ≤ 25 off-source injections are unreliable due to
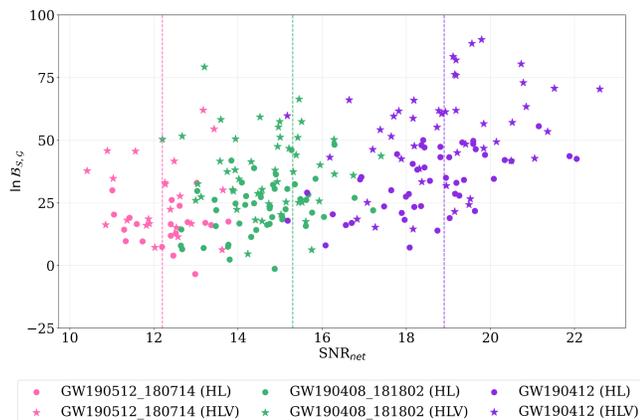
FIG. 2. $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ versus SNR_{net} for IS2 off-source injections of GW190512_180714 (pink), GW190408_181802 (green) and GW190412 (purple). The vertical dashed lines in the respective colors at $SNR_{net} = 12.2, 15.3, 18.9$ indicate the median HLV network match-filter SNRs of the GW events (from Table I). The circles and stars correspond to HL and HLV injections respectively. Gaussian-noise-like events with $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} = -500$ are not shown.

insufficient P_{FA} measurements. Therefore, the four events listed above are excluded from the IS2 analysis. For the remaining 18 GW events, the numbers of injections shown in the last column of Table I include events that are more consistent with Gaussian noise than a GW signal according to BW. As discussed in Sec. III A, these events are assigned $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} = -500$ to indicate low significance.

Figure 2 shows the distribution of off-source injections in IS2 for each GW event in different colors. To avoid clutter, we show only three arbitrarily selected events with contrasting HLV network match-filter SNRs from Table I, namely GW190512_180714 (pink), GW190408_181802 (green) and GW190412 (purple). Each circle (star) data point correspond to an individual HL (HLV) injections. For each GW event in Fig. 2, the off-source injection SNR_{net} are distributed within an approximate range of ± 5 from their respective median HLV network match-filter SNR, indicated by the vertical dashed lines in corresponding colors. The $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ also scales with $\mathcal{I}$, consistent with Ref. [66]. According to Table I, the three events in Fig. 2 also have comparable number of off-source injections in IS2. However, the number of injections for GW190512_180714 (pink) is visibly lower than the other two events, because the plot excludes Gaussian-noiselike events with $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} = -500$. GW190512_180714 has the lowest network match-filter SNR of the three events, so its off-source injections in both the HL and HLV networks also have comparably lower SNR_{net}. Hence, the BW evidences favors the Gaussian noise model more strongly than the signal model for a larger proportion ($\sim 50\%$) of GW190512_180714's offsource injections, cf., ~ 0–5% for the other two events.

The comparison of BW's detection significance (P_{FA}) between the HL and HLV networks is presented in Sec. VI, for all 18 O3 GW events in IS2.

IV. BACKGROUND MEASUREMENTS

In this section, we discuss the suitability of using P_{FA} (as opposed to FAR) as a significance measure for the purpose of our analysis. We then present and discuss the noise background measurements. Using the dataset described in Sec. III A, we obtain the distribution of P_{FA} as a function of $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$.

P_{FA} is the probability that a trigger of a given detection statistic ($\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$) is a false alarm, i.e., nonastrophysical. In the context of hypothesis testing, P_{FA} represents the false positive rate (type I error) and is a dimensionless quantity by definition. In contrast, FAR measures the *temporal* frequency of false alarms producing a detection statistic value equal to or higher than a specified GW candidate event [88]. In other words, FAR is a time-average quantity which conflates BW's performance with engineering factors such as the detector glitch rate. As discussed in Sec. III A, BW is not suitable for a full all-sky search of an observational dataset and is used instead to follow up triggers identified by other burst search pipelines like cWB. In this study, we measure BW's background for the HL and HLV networks using populations of background triggers arbitrarily downselected from the cWB all-sky analysis of the respective O3a time-shifted background data. Since P_{FA} is time-independent and marginalizes over the number of triggers analyzed, it relates directly to how BW is used in this study. It is therefore more appropriate to compare BW's performance between the HL and HLV networks using P_{FA} as a measure of detection significance.[7]

Figure 3 shows the HL and HLV network background as measured by BW with the background trigger datasets described in Sec. III A. P_{FA}, plotted on the vertical axis, is computed as the fraction (i.e. per-trigger probability) of nonastrophysical triggers in the background exceeding the corresponding $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ on the horizontal axis. We restrict the plot to $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} > -20$, the range relevant to real astrophysical signals. Although not shown in Fig. 3, triggers with $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} < -20$ are included in the denominators for computing P_{FA}, that is 1008 and 1134 respectively for HL and HLV. To estimate the uncertainties in our background measurements, we conventionally assume the detector noise background can be modeled as a Poisson process. The shaded regions show the 1σ Poisson uncertainty region for HL and HLV in corresponding colors. In Appendix B,

---

[7]P_{FA} should not be confused with the definition of false alarm probability, $FAP = 1 - \exp(-T_{obs} \times FAR)$ used in other analysis pipelines e.g. PyCBC [89]. FAP is the probability of finding *one or more* noise background events with significance equal to or higher than FAR (of a candidate event) within an observation period T_{obs}.
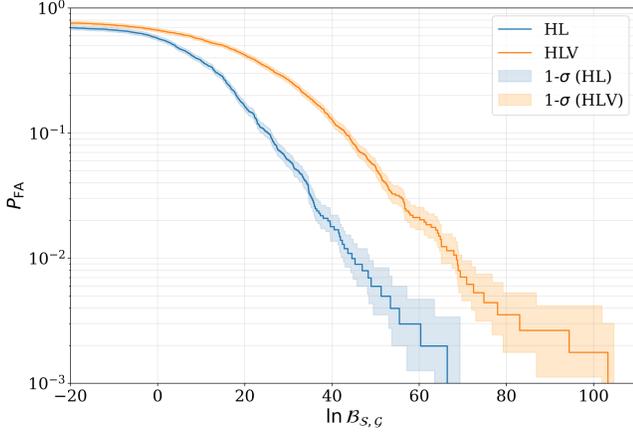
FIG. 3. Background measurements for the BW algorithm. The blue (orange) curve corresponds to the HL (HLV) background measured using the downselected O3a background triggers described in Sec. III A. The shaded bands show the 1-σ Poisson uncertainty regions for each network in corresponding colors.

FIG. 4. Worked example: computing one representative point on the efficiency curve for a significance threshold $P_{FA} = 0.2$. Top: histogram of $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ for the O3a background triggers described in Sec. III A. Bottom: histogram of $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ for IS1. The HL and HLV network histograms are color-coded blue and orange respectively. In both panels, the vertical dashed lines at $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} = 18.1$ (HL) and 34.2 (HLV) indicates the threshold for $P_{FA} = 0.2$. The fraction of injections to the right of the thresholds in the bottom panel yields $P_{det} = 0.74$ (HL) and 0.71 (HLV).

we show the implementation of P_{FA} in Poisson statistics as opposed to FAR, along with the derivation of the Poisson uncertainty regions.

The background measurements show that P_{FA} is higher for HLV than for HL at all $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ as the occurrence of background triggers increases with the number of detectors. As a result, events detected by the HLV network need to attain a higher $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ in order to achieve the same significance (P_{FA}) as the HL network. For example, to achieve $P_{FA} = 0.1$, a HL event requires $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} = 25.6$; cf., $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} = 43.1$ for HLV. Additionally, $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ of the HLV background triggers are higher overall compared to HL. This is because the increased trigger frequency in HLV results in the increased likelihood of coincident triggers which more closely resemble coherent signals, and are therefore recovered with higher detection statistics by BW. Furthermore, the misalignment of the Virgo detector senses a different signal polarization to the two co-aligned LIGO detectors, thus imposing a less stringent constraints on signal coherence. This reduces the efficiency of HLV in discriminating coincident glitches from signals.

V. BAYESWAVE DETECTION EFFICIENCY WITH BBH WAVEFORMS

A. Constructing efficiency curves

In Ref. [44], the performance of a hierarchical pipeline consisting of cWB and BW is quantified using efficiency curves, which show the fraction of injected signal waveforms recovered above various significance thresholds. We use the same approach in this work to study the independent performance of BW. Using IS1 described in Sec. III B 1, we construct efficiency curves for the HL and HLV networks by plotting P_{det} as a function of P_{FA}.

As per Ref. [44], P_{det} is calculated as the fraction of astrophysical events recovered with detection statistic above a threshold. For BW, this threshold is set by the $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ corresponding to a user-selected significance i.e. P_{FA}. As noted in Fig. 3, the $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ threshold is higher for HLV than for HL at a fixed P_{FA}. The following example shows how P_{det} is computed for an arbitrary but representative choice $P_{FA} = 0.2$.

Figure 4 shows histograms of $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ for the HL (blue) and HLV (orange) O3a background triggers in the top panel, and for IS1 in the bottom panel. The $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ thresholds for $P_{FA} = 0.2$ is set by the background triggers in the top panel. In both panels, we indicate the thresholds by the vertical dashed lines at 18.1 (HL, blue) and 34.2 (HLV, orange). With the HL and HLV thresholds established, we turn to the bottom panel of Fig. 4 where we compute P_{det} as the fraction of IS1 injections detected by HL (HLV) greater than the threshold, i.e. to the right of the blue (orange) vertical line. We find $P_{det} = 0.74$ and 0.71 for HL and HLV respectively. The procedure is repeated for P_{FA} in the range $0 \leq P_{FA} \leq 1$ to construct the efficiency curves for HL and HLV.

B. Efficiency analysis

The efficiency curves of IS1 for the characterization of BW's overall burst detection efficiency is shown in Fig. 5. The blue and orange curves correspond to the HL and HLV networks respectively. To indicate the error margins of P_{FA} from the background measurements, we carry over the 1-σ Poisson uncertainty regions onto the horizontal axis of the efficiency curves. From the background measurements, we also noted that the minimum $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ required to achieve a
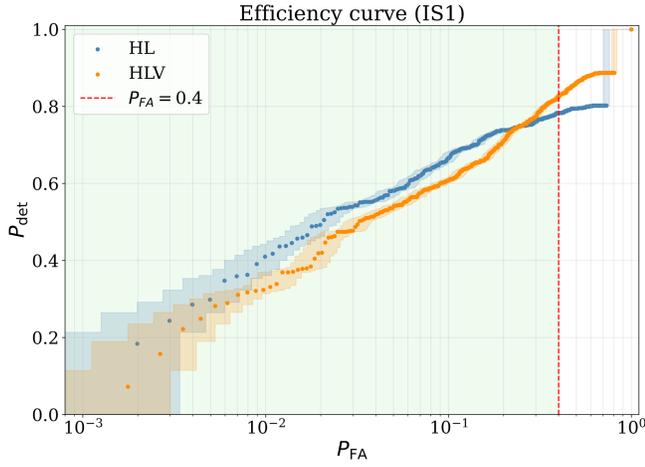
FIG. 5. BW efficiency curves constructed using IS1 for the HL (blue) and HLV (orange) networks. The shaded bands with matching colors are the 1-σ Poisson uncertainty regions for P_{FA}, same as in Fig. 3. The region where $P_{FA} \leq 0.4$ is shaded green to indicate astrophysical relevance.

FIG. 6. Log signal-to-glitch Bayes factor, $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ of the HLV network versus the HL network for IS1. The color bar shows SNR_{net} in HLV for each injection. The diagonal line indicates equal $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ for both networks. The dashed lines at $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}(HL) = 18.1$ and $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}(HLV) = 34.2$ indicate the thresholds for $P_{FA} \leq 0.2$ with the respective networks. Gaussian-noise-like events with $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} = -500$ are excluded in this plot.

given significance reduces with increasing tolerance for P_{FA}. Therefore the efficiency curves show that P_{det} increases with P_{FA}, as more events in IS1 satisfy the reduced $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ threshold. The cluster of data points at $P_{det} = P_{FA} = 1$, disjointed from the rest of the efficiency curves, is an artifact from assigning an arbitrarily low significance of $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} = -500$ to Gaussian-noiselike events. As discussed in Sec. III B 1, these events occupy 20% (11%) of the HL (HLV) IS1 injections. Therefore we observe a discrete jump in the fraction of recovered injections, i.e., P_{det} of HL (HLV) from 0.80 (0.89) to 1. We also note a gap in the P_{FA} between the cluster of data points and the point before $P_{FA} = 1$. This is because the second lowest $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ for the HL and HLV IS1 injections are of order -10^1 according to Fig. 1.

In order to assess the overall detection efficiency of BW with the HL and HLV networks, we focus on the region where P_{FA} is low enough to be astrophysically relevant. We arbitrarily define this region to be where $P_{FA} \leq 0.4$ as indicated by the green shading. In this region, P_{det} of HL is generally higher than HLV, but the opposite is true for $P_{FA} \gtrsim 0.25$. By quantifying the ratio between HL and HLV P_{det} for all data points in $P_{FA} \leq 0.4$, we find that the HL network is only 1.02 times (i.e. 2%) more efficient in detecting IS1 injections than the HLV network on average. Hence, there are no significant differences in BW's overall detection efficiency with a two- or three-detector configuration.

To justify our findings, we show the event-wise comparison of $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ between the HL and HLV networks for IS1 in Fig. 6, color-coded according to the SNR_{net} of HLV.[8] The dashed diagonal line indicates where $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ is equal in

both networks. For a specified detection significance (P_{FA}), the plot can be divided into four quadrants by the corresponding $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ thresholds of the HL and HLV networks. Using $P_{FA} = 0.2$ again as a representative example, we indicate the HL (HLV) threshold with a blue (orange) solid line in Fig. 6. The quadrants classify IS1 events based on their detectability. A successful detection in the HL (HLV) network is when the event $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ is higher than the detection threshold set by the blue (orange) line. By this definition, events in the top left quadrant (shaded orange) are detected by the HLV network only; the bottom right (shaded blue) by the HL network only; the top right by both networks and the bottom left by neither. We note that a fraction of events (in the blue shaded region) are only detected by HL despite having higher $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ in HLV. This is because a successful detection with the HLV network requires the increased $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ to satisfy a higher detection threshold to achieve the same significance as HL. In other words, the advantage of increased $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ in larger detector networks is offset by the higher detection thresholds due to the increased probability of false alarms in the background. This explains why the efficiency curves are comparable between the HL and HLV networks.

From Fig. 6, we can also see that SNR_{net} affects detectability. The top right quadrant contains events with overall higher SNR_{net} compared to the other quadrants. That is, events with higher SNR_{net} and hence higher $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ are more likely to be detected by both HL and HLV. The remaining IS1 events with lower SNR_{net} are distributed across the other three quadrants where they fall short of at least one of the HL or HLV detectability thresholds, as indicated by the orange and blue dashed lines respectively. This is true for all P_{FA}. We discuss the cases where events

---

[8]The HL and HLV network SNR_{net} are equally representative of the ensemble SNR_{net} of IS1 (see Fig. 1). Thus we show only the HLV network SNR_{net} in Fig. 6 to avoid clutter.
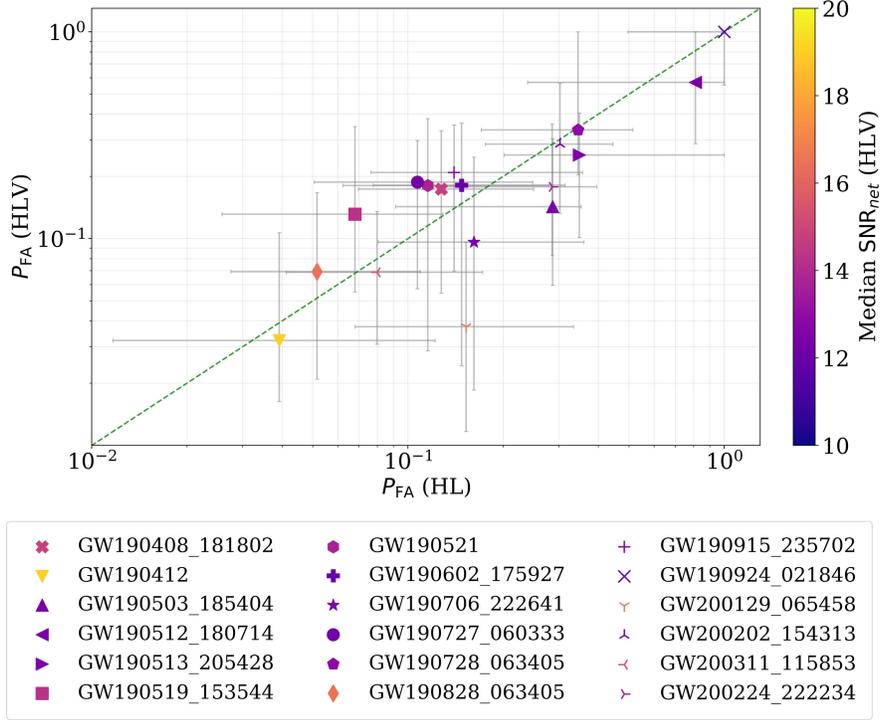
FIG. 7. P_{FA} of the HLV network versus the HL network for O3 GW events in IS2. Each point represents a single GW event as shown in the legend and is color coded by the HLV SNR_{net}. The P_{FA} and SNR_{net} shown are the medians of the off-source injections of the corresponding event; the horizontal (vertical) gray bars span the interquartile range of the HL (HLV) P_{FA} measurements. The diagonal line indicates equal P_{FA} for both networks.

are only detected by one of the two configurations. For events detected only by HLV (orange shaded region), it is straightforward to argue that adding Virgo increases the sensitivity of the network to the signal that are too low to be detected by HL. Consequently, this increases $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ and boosts the detection significance past the required threshold. For the less intuitive case where events are detected only by HL (blue shaded region), we need to justify for two scenarios: (i) where $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ for HLV is higher than HL and (ii) vice versa. The former is discussed above. The latter suggests that the removal of Virgo boosts the signal evidence. This occurs when the addition of Virgo introduces non-Gaussian noise artifacts across the network which outweighs the sensitivity gain for the signal. These effects matter most for low SNR_{net} injections.

In summary, the efficiency curves in Fig. 5 show that BW's overall burst detection performance with the HL and HLV networks are comparable in the nominal astrophysically relevant range $P_{FA} \leq 0.4$. This is because the noisier detector background of the HLV offsets the advantage of increased $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$, as revealed by the granular event analysis in Fig. 6. Additionally we note that for low-SNR_{net} injections at any given significance, adding an extra detector may tip them over or under the detection threshold unpredictably, due to a hard-to-quantify trade-off between the added noise and added sensitivity. High-SNR_{net} injections, on the other hand, are more likely to be detected by both networks.

VI. BAYESWAVE DETECTION SIGNIFICANCE OF O3 GW EVENTS

The analysis with IS1 inferred that the overall detection efficiency of BW is comparable between the HL (two-detector) network and HLV (three-detector) network. Using IS2 described in Sec. III B 2, we conduct a consistency test for the results of IS1 by comparing BW's detection significance of O3 GW events between the two network configurations.

The off-source injections in IS2 correspond to 18 independent O3 GW events. The final column of Table I shows the number of off-source waveforms available for each event. In order to measure the detection significance of these GW events according to BW, we first quantify the significance for each off-source injection. This is done by comparing the recovered $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ in HL and HLV with the corresponding background measurements in Fig. 3. To obtain a single-valued significance measurement for each GW event, we take the median[9] P_{FA} of the corresponding off-source waveforms. Figure 7 shows the median HLV P_{FA} versus that of HL. We use the interquartile range

---

[9]We show the median P_{FA} instead of the mean, because the median value excludes any biases introduced by the Gaussian-noise like events with $P_{FA} = 1$, due to their arbitrarily low detection statistic $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} = -500$.

(IQR), that is the range encompassing the middle 50% of the off-source P_{FA} within each GW event, to represent the uncertainty in our measurements. The horizontal and vertical gray bars show the IQRs for the HL and HLV P_{FA} measurements respectively. We find that all data points are within close proximity of the diagonal line where P_{FA} is equal for HL and HLV. The size of the HL and HLV IQRs are also comparable. This suggests that BW's detection significance of O3 GW events are similar for both networks, further confirming that BW's burst detection performance with the HLV network does not exceed HL when the detector backgrounds are taken into account.

We also note that the median P_{FA} increases with decreasing SNR_{net}, because the colors of the points darkens as one moves from the bottom left to the top right of the plot. According to Fig. 2, GW events with low network match filter SNR have low SNR_{net} off-source injections that are generally more consistent with Gaussian noise. Therefore the median P_{FA} of GW events in the top right of Fig. 7 approaches unity. This observation is consistent with Fig. 6, where events with low SNR_{net} and hence low $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ are only detectable by both HL and HLV when higher P_{FA} are tolerated. Furthermore, the size of the IQRs are within the same order of magnitude as the median P_{FA} viz. the P_{FA} measurement uncertainties are larger for events with lower SNR_{net} off-source injections (top right corner) compared to those with higher SNR_{net} (bottom left corner). The wider IQRs suggest that the increased presence of Gaussian-noiselike injections not only reduces the astrophysical significance, but also increases the uncertainty in the significance measurements for GW events with low network match filer SNR.

Altogether IS2 shows that significance measurements with BW is comparable for the HL and HLV networks, consistent with the findings of the IS1 efficiency curve analysis. We also find that P_{FA} and the uncertainty in its measurement increases with decreasing network match-filter SNR.

VII. CONCLUSION AND DISCUSSION

A. Summary of results

In practice, the source-agnostic BW algorithm is used in conjunction with other search pipelines to enhance detection confidence of GW transients. In this work, however, we study the stand-alone performance of BW with expanded detector networks. Detection confidence of BW is assessed using the algorithm's detection statistics, the log signal-to-glitch Bayes factor $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$, which measures the extent of supporting evidence for the signal model over the glitch model. A previous study shows that $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ increases with increasing number of detectors, $\mathcal{I}$, in a network of GW interferometers [66]. However, the

study did not account for the increase in glitch occurrence and the associated increase in false alarm detections, as more detectors are added to the network. This paper extends Ref. [66] with the goal of determining whether BW's overall burst detection performance is enhanced or reduced as $\mathcal{I}$ increases, when the detector noise background is taken into account. This is done by measuring the noise backgrounds produced by BW and comparing the efficiency curves between the HL (two-detector) and HLV (three-detector) networks.

We obtain the noise backgrounds measurements of BW for the HL and HLV networks by analyzing nonastrophysical triggers, downselected from the cWB analysis of the O3a time-slide background data. The background measurements show that per-trigger false alarm probability P_{FA} is higher in the HLV network than in HL, throughout the astrophysically relevant range $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} \geq -20$. This is due to the increased likelihood of background triggers with an additional detector. We reiterate that the cWB algorithm is only used to downselect triggers for BW's background measurements, we do not investigate cWB's background and/or detection efficiency in this paper.

For the efficiency curve analysis, we implement BW on a population of nonprecessing and nonspinning phenomenological BBH waveforms (IS1) sampled from a parameter space detectable by the Advanced LIGO and Advanced Virgo detectors. IS1 is injected into segments of HL and HLV data spread out across all of O3a, to ensure comparability of the detection statistics with the background measurements. The efficiency curves plots detection efficiency, P_{det}, of IS1 events as a function of the per-trigger false alarm probability, P_{FA}, to characterize BW's performance over a range of significance thresholds. We find similar efficiency curves for the HL and HLV networks within a nominal significance range with plausible astrophysical implications i.e., $P_{FA} \leq 0.4$. In other words, there are no major differences between BW's overall performance with HL and HLV. This counterintuitive finding is justified by event-wise comparison of $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ between the HL and HLV IS1 injections in Fig. 6. The plot reveals that the advantage of increasing $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ with $\mathcal{I}$ is offset by the increased P_{FA}. Adding more detectors to the network increases the likelihood of noise events (i.e. false alarms). Therefore, events in larger detecter networks are required to satisfy higher detection thresholds to achieve the same significance as smaller networks. Additionally, the detectability of events by the HL and HLV networks at any given significance threshold (P_{FA}) scales with SNR_{net}. For events with low SNR_{net}, the $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ and hence P_{FA} in each detector network are more sensitive to subtle changes in detector noise variation. Therefore, the addition of Virgo can unpredictably tip an event over or under the HL or HLV significance threshold.

To check for consistency with the efficiency analysis, we separately analyse a set of O3-like CBC waveforms (IS2), otherwise referred to as off-source injections. Parameters of off-source injections are sampled from the match-filter posteriors of 18 GW events from O3. We use P_{FA} to quantify BW's significance for each GW event. This is evaluated by comparing the $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ of their respective off-source injections with the O3a background measurements. The comparison of P_{FA} between HL and HLV reveals that BW recovers all 18 events with similar significance from both networks. This result is consistent with the IS1 detection efficiency analysis.

Altogether, this study investigates the impact of glitches on the detection significance (P_{FA}) and the overall performance of BW, as a function of $\mathcal{I}$. From two independent analyses with IS1 and IS2, we conclude that there are no significant differences between BW's overall burst detection performance with the HL and HLV networks. Despite the improvement in detection statistic with the addition of Virgo, the associated increase in nonastrophysical background triggers raises the detection statistic threshold which the HLV network need to attain in order to achieve the same per-trigger P_{FA} as HL. Therefore the HLV configuration, despite having more detectors, does not have an advantage over HL in terms of detection efficiency. Our findings are consistent with previous studies [38,90]. Although expanded detector networks improve accuracy of reconstruction and sky localization of the GW signal, Refs. [38] and [90] suggest that HL rejects glitches more efficiently compared to HLV and is therefore preferred in unmodeled burst searches to maximise detection efficiency. This is because HL comprises only of the co-aligned LIGO detectors with similar sensitivities to GW polarization components from all directions, therefore it poses more stringent constraints on signal coherence across the network. On top of that, the overall strain sensitivity of Virgo is lower than the two LIGO detectors in O3a, as shown in Fig. 2 of Ref. [9]. This could be another reason why the larger (HLV) network does not significantly outperform the HL network.

B. Future work

With the recently approved commissioning of LIGO-India with design sensitivity planned to match the LIGO detectors [45], it would be worthwhile for prospective studies on BW's detection efficiency to consider network configurations with three or more detectors of equal sensitivities.

Furthermore, we use trigger lists generated by the cWB algorithm from the LVK O3 all-sky burst search [38] to downselect triggers for BW background measurements in this study. However, Ref. [90] conducted the same search using the cWB algorithm enhanced by machine-learning

(ML) which shows improved overall search sensitivity compared to the standard cWB. We therefore suggest a complementary study to follow-up on whether the BW background measurements can be improved if the triggers are downselected from the ML-enhanced cWB trigger list instead.

While BW targets a broad range of unmodeled GW bursts, this study considers only CBC waveforms as they are the only source category detected in the LVK observing runs to date. One can generalize this study to alternative transient sources like supernovae and generic white noise bursts, but the analysis presented in this work is limited to comparing the overall trends of BW's independent performance between the HL and HLV networks. We did not study the sensitivity of BW to specific types of burst signal because BW is not used independently in practice, but rather to follow-up cWB triggers to enhance detection confidence. With promising outlooks for the ML-enhanced cWB [90] and O4 in progress, future work should consider assessing the joint performance of the ML-enhanced cWB algorithm with BW for different types of burst sources as in Ref. [44].

ACKNOWLEDGMENTS

APPENDIX A: BW CONFIGURATION

To assist with reproducibility, we detail the BW settings for the background measurements and injection analyses. The following settings are adapted from the BW analysis used in the O3 all-sky burst search [38].

To down-select candidates from the cWB trigger list for BW's background measurements, we specify the significance threshold, $\rho_{threshold} = 7$, as a first cut. We further reduce the dataset by keeping only a fraction of triggers satisfying $\rho > \rho_{threshold}$. This fraction is denoted by X in the main text.

For the signal injections, we use the O3a calibrated strain data for the LIGO Hanford (H1), LIGO Livingston (L1) and Virgo (V1) detectors [85]. The frame calibration includes a noise subtraction procedure detailed in Ref. [91]. The Advanced LIGO (H1 and L1) noise subtraction targets noise from beam jitter, detector calibration lines and the main power grid line (at 60 Hz) [61]. For Advanced Virgo, we use low-latency (online) strain data which includes subtraction of frequency noise from the input laser, Michelson noise from displacement of the beam splitter mirrors, amplitude noise from auxiliary modulation and scattered light noise [92].

For all analyses, we set the low frequency cutoff at 20 Hz by convention [93]. The sampling rate is set at 2048 Hz to achieve a Nyquist frequency of 1024 Hz. For PSD estimation, i.e., to construct the model $\mathcal{N}$, we employ the *BayesLine* algorithm. The BW analysis segment length is set to 4 seconds, even though GW burst signals (especially CBCs) are typically shorter. This is to ensure that detector noise is relatively stationary in analysis segment for accurate prediction of the noise spectral density with *BayesLine*. Altogether, our search targets GW bursts signals with duration of milliseconds up to a few seconds, with frequencies in the 20–1024 Hz frequency band of Advanced LIGO and Advanced Virgo at O3a sensitivities.

APPENDIX B: POISSON NOISE BACKGROUND

The Poisson process models a series of randomly occurring events where the average time between events are known, but not the exact time of arrival of each event. Events modeled as Poisson process are expected to have a probability mass function given by

$$P(n,\lambda) = \frac{\lambda^n \exp^{-\lambda}}{n!}. \qquad (B1)$$

Otherwise known as the Poisson distribution, Equation (B1) measures the probability P of n number of events occurring within a population for a given rate parameter, $\lambda > 0$. In this context, "population" refers to a group of events in a fixed temporal or spatial interval. By definition, λ is the expected number of events in a given population, independent of the type of interval specified i.e. it is dimensionless.

1. P_{FA} vs. FAR in modeling Poisson noise

The noise background of the Advanced LIGO and Virgo detectors are modeled as a Poisson process in the standard LVK GW transient searches [8,9,11]. In modeling a Poisson noise background, P_{FA} and FAR play an analogous role of representing the rate of noise events, which directly influences the rate parameter λ. In the case of FAR where rate is measured in units of time, the time of observation

T_{obs} is the interval required to obtain the expected number of noise events in the background, $\lambda = T_{\mathrm{obs}} \times \mathrm{FAR}$. Conversely, P_{FA} measures the noise occurrence rate in units of events. Therefore $\lambda = N_{\mathrm{obs}} \times \mathrm{FAR}$, where the interval is now given by the total number of events observed N_{obs}. One can then show the relationship between P_{FA} and FAR as $P_{\mathrm{FA}} = (N_{\mathrm{obs}}/T_{\mathrm{obs}}) \times \mathrm{FAR}$.

2. Poisson uncertainty regions of P_{FA}

Since the background triggers used for BW's background measurements in Sec. IV are subsets of the cWB all-sky analysis of the full O3a time-shifted background of the LIGO-Virgo network, we can thereby assume the triggers obey the Poisson distribution. Consequently, we can use the standard deviation (σ) of Eq. (B1) to represent the error margins of our P_{FA} measurements. We show the derivation as follows.

In the background measurements shown in Fig. 3, P_{FA} (on the vertical axis) is computed as the fraction of background triggers recovered by BW with $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}$ exceeding the corresponding threshold, $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}^*$ (on the horizontal axis) viz.

$$P_{\mathrm{FA}} = \frac{n(\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} \geq \ln \mathcal{B}_{\mathcal{S},\mathcal{G}}^*)}{n_{\mathrm{tot}}}, \qquad (B2)$$

where n_{tot} is the total number of triggers in the background dataset (the population). The numerator is essentially the expected occurrence of events exceeding $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}^*$, hence $\lambda = n(\ln \mathcal{B}_{\mathcal{S},\mathcal{G}} \geq \ln \mathcal{B}_{\mathcal{S},\mathcal{G}}^*)$. One can then derive the 1-σ error margin for counting the number of events n exceeding $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}^*$ from the variance of the Poisson distribution:

$$\sigma = \sqrt{\sum_{n=1}^{\lambda}(n-\lambda)^2 P(n,\lambda)} = \sqrt{\lambda}. \qquad (B3)$$

Combining Eqs. (B2) and (B3), the 1σ Poisson uncertainty region of P_{FA} for a given $\ln \mathcal{B}_{\mathcal{S},\mathcal{G}}^*$ is bounded by

$$\frac{\lambda - \sqrt{\lambda}}{n_{\mathrm{tot}}} \leq P_{\mathrm{FA}} \leq \frac{\lambda + \sqrt{\lambda}}{n_{\mathrm{tot}}}, \qquad (B4)$$

as indicated by the shaded regions in Figs. 3 and 5.

To check for viability, we plot the cumulative number of triggers against P_{FA} in Fig. 8 and the shaded regions show the 1-, 2- and 3-σ P_{FA} Poisson uncertainty regions. We compare our plots to the O3 backgrounds in Ref. [38] measured with inverse FAR. Even though we use a difference quantity to measure significance, the relative extent of the shaded regions are comparable. It is therefore appropriate to use the Poisson uncertainty described above as the error margins for our P_{FA} measurements.
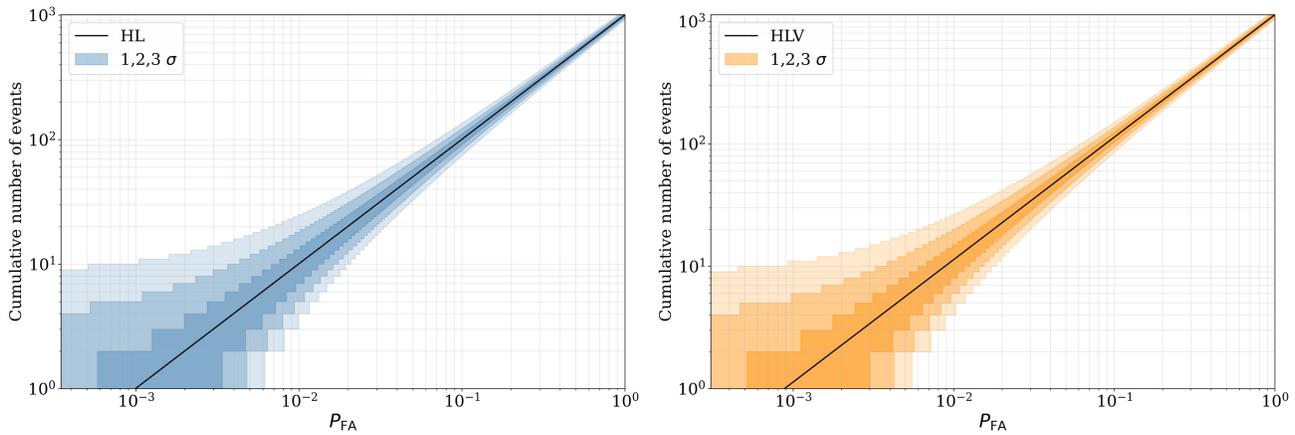
FIG. 8. Cumulative number of events versus per-trigger false alarm probability, P_{FA} for the BW background measurements of the HL (left, 1008 triggers) and HLV (right, 1134 triggers) network.

[1] J. Aasi *et al.* (LIGO Scientific Collaboration), Advanced LIGO, Classical Quantum Gravity **32,** 074001 (2015).

[2] F. Acernese *et al.* (Virgo Collaboration), Advanced Virgo: A second-generation interferometric gravitational wave detector, Classical Quantum Gravity **32,** 024001 (2015).

[3] T. Akutsu *et al.* (KAGRA Collaboration), Overview of KAGRA: Detector design and construction history, Prog. Theor. Exp. Phys. **2021,** 05A101 (2021).

[4] Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto (KAGRA Collaboration), Interferometer design of the KAGRA gravitational wave detector, Phys. Rev. D **88,** 043007 (2013).

[5] K. Somiya, Detector configuration of KAGRA-the Japanese cryogenic gravitational-wave detector, Classical Quantum Gravity **29,** 124007 (2012).

[6] R. Abbott *et al.*, First joint observation by the underground gravitational-wave detector KAGRA with GEO 600, Prog. Theor. Exp. Phys. **2022,** 063F01 (2022).

[7] K. L. Dooley, J. R. Leong, T. Adams, C. Affeldt, A. Bisht, C. Bogan, J. Degallaix, C. Gräf, S. Hild, J. Hough *et al.*, GEO 600 and the GEO-HF upgrade program: Successes and challenges, Classical Quantum Gravity **33,** 075009 (2016).

[8] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs, Phys. Rev. X **9,** 031040 (2019).

[9] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), GWTC-2: Compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run, Phys. Rev. X **11,** 021053 (2021).

[10] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo

during the first half of the third observing run, Phys. Rev. D **109,** 022001 (2024).

[11] R. Abbott *et al.* (KAGRA Collaboration, LIGO Scientific Collaboration, and Virgo Collaboration), GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, Phys. Rev. X **13,** 041039 (2023).

[12] C. L. Fryer and K. C. B. New, Gravitational waves from gravitational collapse, Living Rev. Relativity **14,** 1 (2011).

[13] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Optically targeted search for gravitational waves emitted by core-collapse supernovae during the first and second observing runs of Advanced LIGO and Advanced Virgo, Phys. Rev. D **101,** 084002 (2020).

[14] E. Stopnitzky and S. Profumo, Gravitational waves from gamma-ray pulsar glitches, Astrophys. J. **787,** 114 (2014).

[15] S. Mereghetti, J. A. Pons, and A. Melatos, Magnetars: Properties, origin and evolution, Space Sci. Rev. **191,** 315 (2015).

[16] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Search for transient gravitational-wave signals associated with magnetar bursts during Advanced LIGO's second observing run, Astrophys. J. **874,** 163 (2019).

[17] M. Ebersold and S. Tiwari, Search for nonlinear memory from subsolar mass compact binary mergers, Phys. Rev. D **101,** 104041 (2020).

[18] M. Sakellariadou, Gravitational waves emitted from infinite strings, Phys. Rev. D **42,** 354 (1990).

[19] T. Damour and A. Vilenkin, Gravitational radiation from cosmic (super)strings: Bursts, stochastic background, and observational windows, Phys. Rev. D **71,** 063510 (2005).

[20] R. Abbott *et al.* (KAGRA Collaboration, LIGO Scientific Collaboration, and Virgo Collaboration), Constraints on cosmic strings using data from the third Advanced

LIGO-Virgo observing run, Phys. Rev. Lett. **126,** 241102 (2021).

[21] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries, Phys. Rev. D **85,** 122006 (2012).

[22] D. A. Brown, I. Harry, A. Lundgren, and A. H. Nitz, Detecting binary neutron star systems with spin in advanced gravitational-wave detectors, Phys. Rev. D **86,** 084017 (2012).

[23] P. Ajith, N. Fotopoulos, S. Privitera, A. Neunzert, N. Mazumder, and A. J. Weinstein, Effectual template bank for the detection of gravitational waves from inspiralling compact binaries with generic spins, Phys. Rev. D **89,** 084041 (2014).

[24] T. Dal Canton and I. W. Harry, Designing a template bank to observe compact binary coalescences in Advanced LIGO's second observing run, arXiv:1705.01845.

[25] A. Taracchini, A. Buonanno, Y. Pan, T. Hinderer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, A. H. Mroué, H. P. Pfeiffer, M. A. Scheel, B. Szilágyi, N. W. Taylor, and A. Zenginoglu, Effective-one-body model for black-hole binaries with generic mass ratios and spins, Phys. Rev. D **89,** 061502 (2014).

[26] L. Blanchet, Gravitational radiation from Post-Newtonian sources and inspiralling compact binaries, Living Rev. Relativity **17,** 2 (2014).

[27] S. Privitera, S. R. P. Mohapatra, P. Ajith, K. Cannon, N. Fotopoulos, M. A. Frei, C. Hanna, A. J. Weinstein, and J. T. Whelan, Improving the sensitivity of a search for coalescing binary black holes with nonprecessing spins in gravitational wave data, Phys. Rev. D **89,** 024003 (2014).

[28] S. V. Dhurandhar and B. S. Sathyaprakash, Choice of filters for the detection of gravitational waves from coalescing binaries. II. Detection in colored noise, Phys. Rev. D **49,** 1707 (1994).

[29] R. Lynch, S. Vitale, R. Essick, E. Katsavounidis, and F. Robinet, Information-theoretic approach to the gravitational-wave burst detection problem, Phys. Rev. D **95,** 104046 (2017).

[30] P. J. Sutton, G. Jones, S. Chatterji, P. Kalmus, I. Leonor, S. Poprocki, J. Rollins, A. Searle, L. Stein, M. Tinto, and M. Was, X-Pipeline: An analysis package for autonomous gravitational-wave burst searches, New J. Phys. **12,** 053034 (2010).

[31] V. Skliris, M. R. K. Norman, and P. J. Sutton, Real-time detection of unmodelled gravitational-wave transients using convolutional neural networks, arXiv:2009.14611.

[32] S. Klimenko, I. Yakushin, A. Mercer, and G. Mitselmakher, A coherent method for detection of gravitational wave bursts, Classical Quantum Gravity **25,** 114029 (2008).

[33] S. Klimenko, G. Vedovato, M. Drago, F. Salemi, V. Tiwari, G. A. Prodi, C. Lazzaro, K. Ackley, S. Tiwari, C. F. Da Silva, and G. Mitselmakher, Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors, Phys. Rev. D **93,** 042004 (2016).

[34] M. Drago, S. Klimenko, C. Lazzaro, E. Milotti, G. Mitselmakher, V. Necula, B. O'Brian, G. A. Prodi, F. Salemi, M. Szczepanczyk, S. Tiwari, V. Tiwari,

V. Gayathri, G. Vedovato, and I. Yakushin, Coherent WaveBurst a pipeline for unmodeled gravitational-wave data analysis, SoftwareX **14,** 100678 (2021).

[35] S. Klimenko, Wavescan: Multiresolution regression of gravitational-wave data, arXiv:2201.01096.

[36] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), All-sky search for short gravitational-wave bursts in the first Advanced LIGO run, Phys. Rev. D **95,** 042003 (2017).

[37] B. P. Abbott et al. (LIGO Scientific Collaboration and Virgo Collaboration), All-sky search for short gravitational-wave bursts in the second Advanced LIGO and Advanced Virgo run, Phys. Rev. D **100,** 024017 (2019).

[38] R. Abbott et al. (KAGRA Collaboration, LIGO Scientific Collaboration, and Virgo Collaboration), All-sky search for short gravitational-wave bursts in the third Advanced LIGO and Advanced Virgo run, Phys. Rev. D **104,** 122004 (2021).

[39] T. B. Littenberg and N. J. Cornish, Bayesian approach to the detection problem in gravitational wave astronomy, Phys. Rev. D **80,** 063007 (2009).

[40] T. B. Littenberg and N. J. Cornish, Bayesian inference for spectral estimation of gravitational wave detector noise, Phys. Rev. D **91,** 084034 (2015).

[41] N. J. Cornish and T. B. Littenberg, Bayeswave: Bayesian inference for gravitational wave bursts and instrument glitches, Classical Quantum Gravity **32,** 135012 (2015).

[42] N. J. Cornish, T. B. Littenberg, B. Bécsy, K. Chatziioannou, J. A. Clark, S. Ghonge, and M. Millhouse, BayesWave analysis pipeline in the era of gravitational wave observations, Phys. Rev. D **103,** 044006 (2021).

[43] BayesWave source code repository, https://git.ligo.org/lscsoft/bayeswave/.

[44] J. B. Kanner, T. B. Littenberg, N. Cornish, M. Millhouse, E. Xhakaj, F. Salemi, M. Drago, G. Vedovato, and S. Klimenko, Leveraging waveform complexity for confident detection of gravitational waves, Phys. Rev. D **93,** 022002 (2016).

[45] B. Iyer, T. Souradeep, C. S. Unnikrishnan, S. Dhurandhar, S. Raja, and A. Sengupta, Technical Report No. LIGO-M1100296, https://dcc.ligo.org/LIGO-M1100296/public (2011).

[46] B. P. Abbott et al. (KAGRA Collaboration, LIGO Scientific Collaboration, and Virgo Collaboration), Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA, Living Rev. Relativity **21,** 3 (2018).

[47] B. F. Schutz, Networks of gravitational wave detectors and three figures of merit, Classical Quantum Gravity **28,** 125023 (2011).

[48] D. Davis, T. B. Littenberg, I. M. Romero-Shaw, M. Millhouse, J. McIver, F. Di Renzo, and G. Ashton, Subtracting glitches from gravitational-wave detector data during the third LIGO-Virgo observing run, Classical Quantum Gravity **39,** 245013 (2022).

[49] M. Zevin, S. Coughlin, S. Bahaadini, E. Besler et al., Gravity Spy: Integrating Advanced LIGO detector characterization, machine learning, and citizen science, Classical Quantum Gravity **34,** 064003 (2017).

[50] J. Powell, Parameter estimation and model selection of gravitational wave signals contaminated by transient detector noise glitches, Classical Quantum Gravity **35,** 155017 (2018).

[51] F. Robinet, N. Arnaud, N. Leroy, A. Lundgren, D. Macleod, and J. McIver, Omicron: A tool to characterize transient noise in gravitational-wave detectors, SoftwareX **12,** 100620 (2020).

[52] R. E. Colgan, K. R. Corley, Y. Lau, I. Bartos, J. N. Wright, Z. Márka, and S. Márka, Efficient gravitational-wave glitch identification from environmental data through machine learning, Phys. Rev. D **101,** 102003 (2020).

[53] J. Glanzer, S. Banagiri, S. B. Coughlin, S. Soni, M. Zevin, C. P. L. Berry, O. Patane, S. Bahaadini, N. Rohani, K. Crowston, V. Kalogera, C. Østerlund, L. Trouille, and A. Katsaggelos, Data quality up to the third observing run of Advanced LIGO: Gravity Spy glitch classifications, Classical Quantum Gravity **40,** 065004 (2023).

[54] M. Cabero, A. Lundgren, A. H. Nitz, T. Dent, D. Barker, E. Goetz, J. S. Kissel, L. K. Nuttall, P. Schale, R. Schofield, and D. Davis, Blip glitches in Advanced LIGO data, Classical Quantum Gravity **36,** 155010 (2019).

[55] L. K. Nuttall, T. J. Massinger, J. Areeda, J. Betzwieser, S. Dwyer, A. Effler, R. P. Fisher, P. Fritschel, J. S. Kissel, A. P. Lundgren, D. M. Macleod, D. Martynov, J. McIver, A. Mullavey, D. Sigg, J. R. Smith, G. Vajente, A. R. Williamson, and C. C. Wipf, Improving the data quality of Advanced LIGO based on early engineering run results, Classical Quantum Gravity **32,** 245005 (2015).

[56] S. Soni, C. Austin, A. Effler *et al.* (LIGO Scientific Collaboration), Reducing scattered light in LIGO's third observing run, Classical Quantum Gravity **38,** 025016 (2021).

[57] R. DeRosa, J. C. Driggers, D. Atkinson, H. Miao, V. Frolov, M. Landry, J. A. Giaime, and R. X. Adhikari, Global feedforward vibration isolation in a km scale interferometer, Classical Quantum Gravity **29,** 215008 (2012).

[58] V. Tiwari, M. Drago, V. Frolov, S. Klimenko, G. Mitselmakher, V. Necula, G. Prodi, V. Re, F. Salemi, G. Vedovato, and I. Yakushin, Regression of environmental noise in LIGO data, Classical Quantum Gravity **32,** 165014 (2015).

[59] G. D. Meadors, K. Kawabe, and K. Riles, Increasing LIGO sensitivity by feedforward subtraction of auxiliary length control noise, Classical Quantum Gravity **31,** 105014 (2014).

[60] J. C. Driggers, S. Vitale, A. P. Lundgren *et al.*, Improving astrophysical parameter estimation via offline noise subtraction for Advanced LIGO, Phys. Rev. D **99,** 042001 (2019).

[61] D. Davis, T. Massinger, A. Lundgren, J. C. Driggers, A. L. Urban, and L. Nuttall, Improving the sensitivity of Advanced LIGO using noise subtraction, Classical Quantum Gravity **36,** 055011 (2019).

[62] G. Vajente, Y. Huang, M. Isi, J. C. Driggers, J. S. Kissel, M. J. Szczepańczyk, and S. Vitale, Machine-learning nonstationary noise out of gravitational-wave detectors, Phys. Rev. D **101,** 042003 (2020).

[63] R. Ormiston, T. Nguyen, M. Coughlin, R. X. Adhikari, and E. Katsavounidis, Noise reduction in gravitational-wave data via deep learning, Phys. Rev. Res. **2,** 033066 (2020).

[64] D. Davis and M. Walker, Detector characterization and mitigation of noise in ground-based gravitational-wave interferometers, Galaxies **10,** 12 (2022).

[65] T. B. Littenberg, J. B. Kanner, N. J. Cornish, and M. Millhouse, Enabling high confidence detections of gravitational-wave bursts, Phys. Rev. D **94,** 044050 (2016).

[66] Y. S. C. Lee, M. Millhouse, and A. Melatos, Enhancing the gravitational-wave burst detection confidence in expanded detector networks with the BayesWave pipeline, Phys. Rev. D **103,** 062002 (2021).

[67] B. Bécsy, P. Raffai, N. J. Cornish, R. Essick, J. Kanner, E. Katsavounidis, T. B. Littenberg, M. Millhouse, and S. Vitale, Parameter estimation for gravitational-wave bursts with the BayesWave pipeline, Astrophys. J. **839,** 15 (2017).

[68] S. Ghonge, K. Chatziioannou, J. A. Clark, T. Littenberg, M. Millhouse, L. Cadonati, and N. Cornish, Reconstructing gravitational wave signals from binary black hole mergers with minimal assumptions, Phys. Rev. D **102,** 064056 (2020).

[69] P. J. Green, Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination, Biometrika **82,** 711 (1995).

[70] W. Anderson, P. Brady, D. Chin, J. Creighton, K. Riles, and J. Whelan, Beam pattern response functions and times of arrival for earthbound interferometer, LIGO Internal Document, Report No. LIGOT010110-00-Z, 2011, https://dcc.ligo.org/public/0012/T010110/001/T010110.pdf.

[71] J. Wang and R. H. Swendsen, Replica Monte Carlo simulation (Revisited), Prog. Theor. Phys. Suppl. **157,** 317 (2005).

[72] P. Goggans and C. Chan, Bayesian inference and methods in science and engineering, Proceedings of the 29th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Oxford, Mississippi (American Institute of Physics, USA, 2004), https://inspirehep.net/authors/2421297.

[73] L. Blackburn, L. Cadonati *et al.*, The LSC glitch group: Monitoring noise transients during the fifth LIGO science run, Classical Quantum Gravity **25,** 184004 (2008).

[74] B. P. Abbott, R. Abbott *et al.*, Search for gravitational-wave bursts in the first year of the fifth LIGO science run, Phys. Rev. D **80,** 102001 (2009).

[75] L. S. Collaboration, Ligo: The laser interferometer gravitational-wave observatory, Rep. Prog. Phys. **72,** 076901 (2009).

[76] J. Aasi, J. Abadie *et al.*, The characterization of Virgo data and its impact on gravitational-wave searches, Classical Quantum Gravity **29,** 155002 (2012).

[77] J. Abadie *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), All-sky search for gravitational-wave bursts in the second joint ligo-virgo run, Phys. Rev. D **85,** 122007 (2012).

[78] M. Wąs, M.-A. Bizouard, V. Brisson, F. Cavalier, M. Davier, P. Hello, N. Leroy, F. Robinet, and M. Vavoulidis, On the background estimation by time slides in a network of gravitational wave detectors, Classical Quantum Gravity **27,** 015005 (2010).

[79] J. Abadie, B. P. Abbott *et al.*, All-sky search for gravitational-wave bursts in the first joint LIGO-GEO-Virgo run, Phys. Rev. D **81,** 102001 (2010).

[80] B. P. Abbott, R. Abbott *et al.*, Observing gravitational-wave transient GW150914 with minimal assumptions, Phys. Rev. D **93,** 122004 (2016).

[81] S. Klimenko, G. Vedovato, V. Necula, F. Salemi, M. Drago, E. Chassande-Mottin, V. Tiwari, C. Lazzaro, B. O'Brian, M. Szczepanczyk, S. Tiwari, and V. Gayathri, CWB pipeline library: 6.4.0 (2021), 10.5281/zenodo.4419902.

[82] D. Lopez, V. Gayathri, A. Pai, I. S. Heng, C. Messenger, and S. K. Gupta, Utilizing Gaussian mixture models in all-sky searches for short-duration gravitational wave bursts, Phys. Rev. D **105,** 063024 (2022).

[83] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal, Phys. Rev. D **93,** 044006 (2016).

[84] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era, Phys. Rev. D **93,** 044007 (2016).

[85] R. Abbott *et al.* (KAGRA Collaboration, LIGO Scientific Collaboration, and Virgo Collaboration), Open data from the third observing run of LIGO, Virgo, KAGRA and GEO, Astrophys. J. Suppl. Ser. **267,** 29 (2023).

[86] LIGO Scientific Collaboration and Virgo Collaboration, LIGO Virgo strain data from observing run O3a, https:// gwosc.org/O3/O3a/ (2021).

[87] L. S. Finn, Detection, measurement, and gravitational radiation, Phys. Rev. D **46,** 5236 (1992).

[88] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), GW150914: First results from the search for binary black hole coalescence with Advanced LIGO, Phys. Rev. D **93,** 122003 (2016).

[89] S. A. Usman, A. H. Nitz, I. W. Harry, C. M. Biwer, D. A. Brown, M. Cabero, C. D. Capano, T. Dal Canton, T. Dent, S. Fairhurst *et al.*, The PyCBC search for gravitational waves from compact binary coalescence, Classical Quantum Gravity **33,** 215004 (2016).

[90] M. J. Szczepańczyk, F. Salemi, S. Bini, T. Mishra, G. Vedovato, V. Gayathri, I. Bartos, S. Bhaumik, M. Drago, O. Halim, C. Lazzaro, A. Miani, E. Milotti, G. A. Prodi, S. Tiwari, and S. Klimenko, Search for gravitational-wave bursts in the third Advanced LIGO-Virgo run with coherent WaveBurst enhanced by machine learning, Phys. Rev. D **107,** 062002 (2023).

[91] G. Vajente, Y. Huang, M. Isi, J. C. Driggers, J. S. Kissel, M. J. Szczepańczyk, and S. Vitale, Machine-learning nonstationary noise out of gravitational-wave detectors, Phys. Rev. D **101,** 042003 (2020).

[92] F. Acernese *et al.* (Virgo Collaboration), Calibration of Advanced Virgo and reconstruction of the detector strain $h(t)$ during the observing run O3, Classical Quantum Gravity **39,** 045006 (2022).

[93] B. P. Abbott, R. Abbott, T. D. Abbott *et al.*, A guide to LIGO-Virgo detector noise and extraction of transient gravitational-wave signals, Classical Quantum Gravity **37,** 055002 (2020).