Obtaining statistical significance of gravitational wave signals in hierarchical search

Kanchan Soni[*], Sanjeev Dhurandhar,[†] and Sanjit Mitra[‡]

Inter-University Centre for Astronomy and Astrophysics, Pune 411007, India

Gravitational wave (GW) astronomy has experienced remarkable growth in recent years, driven by advancements in ground-based detectors. While detecting compact binary coalescences (CBCs) has become routine, searching for more complex ones, such as mergers involving eccentric and precessing binaries and subsolar mass binaries, has presented persistent challenges. These challenges arise from using the standard matched filtering algorithm, whose computational cost increases with the dimensionality and size of the template bank. This urges the pressing need for faster search pipelines to efficiently identify GW signals, leading to the emergence of the hierarchical search strategy that reduces the computational cost of matched filtering in the search. This method looks for potential candidate events using a coarse bank of templates (with reduced density and sampling rate) in the first stage, which are then followed up in the second stage with the usual template bank (with optimal density and sampling rate) but only in the neighborhood of the parameter space of the candidate events identified in the first stage. Although the hierarchical search was demonstrated to speed up the standard PyCBC analysis by more than a factor of 20 in a previous work outlined in Soni *et al.* [Phys. Rev. D **105**, 064005 (2022)], assigning statistical significance to detected signals, especially in the presence of non-Gaussian noise, was done in a heuristic way. This article introduces a method for background estimation in a two-stage hierarchical search. Our method models the distribution of background triggers obtained from time-shifted triggers in a two-detector network. This modeling precisely aligns with the background distribution across critical signal-to-noise ratios (SNRs), where distinguishing between signal and noise is particularly challenging. It incorporates a fitting procedure to extrapolate to higher detection statistic values. Using this background estimation approach, we assessed the sensitivity of the hierarchical search compared to the standard PyCBC search. Our findings indicate that the hierarchical search attains a sensitive volume-time product comparable to the standard two-detector PyCBC search. This equivalence holds for an inverse false alarm rate of 10 years by a factor ranging from 0.99 ± 0.144 and 1.02 ± 0.141 within the chirp mass range of approximately 1.4 to $20 M_\odot$. Our methodology accomplishes this while substantially reducing the overall computational cost of the analysis. Specifically, our pipeline exhibits a remarkable speed-up, nearly 13 times faster than PyCBC analysis, including background estimation.

I. INTRODUCTION

Gravitational wave (GW) astronomy has experienced impressive development since the landmark discovery of the first GW signal from the merger of the binary black hole (BBH), GW150914 [1]. Since then, the LIGO Scientific, Virgo, and KAGRA (LVK) collaboration has compiled an impressive catalog of nearly 90 GW sources [2] comprising numerous binary black holes, two binary neutron stars (BNS) [3,4], and two neutron star-black holes (NSBH) [5]. A majority of these detections have been possible through offline search pipelines [6–13], which primarily use matched filtering [14–17], a model-based search technique that cross-correlates data from Advanced LIGO [18] and Advanced Virgo [19] with a bank of modeled signals called *templates*.

In the current state-of-the-art, the search is mostly conducted for quadrupolar GW signals originating from mergers of compact binaries with quasicircular orbits. With recent advancements in ground-based GW detectors, the ongoing upgrades in KAGRA [20], and the establishment of proposed detectors such as LIGO-Aundha [21,22] and third-generation detectors [23], the future holds promise for many more discoveries. It may even become possible to detect GW signals from binary systems, where parameter spaces expand by several orders of magnitude. This includes scenarios such as binary systems with orbital precession [24] and eccentricity, as well as cases where the

[*]kanchansoni@iucaa.in
[†]sanjeev@iucaa.in
[‡]sanjit@iucaa.in

density of templates for spin-aligned systems increases, such as in subsolar mass compact binaries.

Currently, such searches are either deferred or conducted within highly restrictive parameter spaces owing to computational limitations. This limitation arises because of the computational cost of matched filtering that scales with the template length, number of templates used, and dimensionality of the parameter space searched in the year-observed data. Consequently, there is an urgent demand for faster search pipelines that efficiently identify GW signals using these highly sensitive detectors. In this context, the application of hierarchical search methodology [25,26] has emerged as a promising approach for detecting GW signals.

The hierarchical search method uses multiple banks of varying densities for matched filtering detector's data. Initially proposed by Mohanty and Dhurandhar [27], matched filtering was performed by hierarchically searching over the chirp mass of nonspinning binary systems using Newtonian waveforms in stationary Gaussian noise. This approach was extended in subsequent studies to incorporate hierarchy over component masses using post-Newtonian (1.5PN) waveforms in Mohanty [28] and over three parameters, component masses, and time of coalescence (t_c), in Sengupta et al. [29,30] using second-order post-Newtonian waveforms. It was also discovered that reducing the data sampling rate during the initial stage of the hierarchy improves the performance.

Gadre et al. [25] achieved a significant improvement in the hierarchical search method. Their work expanded the methodology to include all intrinsic parameters, such as binary component masses and spins, and introduced a two-detector coincidence analysis into the algorithm. This advancement resulted in a remarkable speed-up of more than one order of magnitude in the Gaussian noise compared to traditional analysis techniques. Further exploration and development of this approach were conducted in Soni et al. [31], where a search pipeline was established and applied to data from the first two observing runs (O1 and O2) of Advanced LIGO. The study demonstrated the effectiveness of the hierarchical search method, which could detect GW events previously identified by the PyCBC search and reported in Abbott et al. [32]. Notably, the hierarchical search significantly reduced the matched filtering computation by approximately 20-fold compared to the standard PyCBC (flat) search method.

The hierarchical search methodology described in Soni et al. [31] involved a two-stage matched filtering search within the PyCBC framework. In the first stage, a coarse search is performed on the data sampled at a low frequency using a sparsely sampled template bank called coarse bank, generated using the hybrid geometric-random method [33,34]. This enables faster matched-filtering operations, thereby facilitating a rapid parameter space scan. The objective of the coarse search is to identify potential coincident triggers that may be GW signals for the detector network. In the second stage, a more refined search is carried out in the vicinity, or neighborhood (nbhd), of the coincident trigger's parameter space identified during the coarse search. A nbhd of a coarse template is a region in parameter space where the match of any template with itself lies above 0.75. This step allows matched filtering data with only a few tens or hundreds of templates in the nbhd region. At this stage, matched filtering is executed at a higher sampling rate, enabling a meticulous investigation within a limited region of the parameter space. This approach maximizes the likelihood of detecting true GW signals, as the finer search focuses on regions of the parameter space where these signals are more likely to manifest.

The two-stage hierarchical search strategy significantly reduces the computational cost of matched filtering. However, accurately assigning statistical significance to detected signals, particularly in real data containing non-Gaussian noise transients, remains challenging for this method. Specifically, estimating a GW event's false alarm rate (FAR) is not straightforward and requires careful consideration.

The statistical significance of a GW event detected in one or more detectors is described by its FAR, which measures the likelihood of the event being a noise event rather than a true GW signal. There are various methods to estimate the FAR of a detected candidate. One commonly used approach involves artificially introducing time shifts to the data from one detector and searching for triggers coincident in time and the template parameters with the data from another detector [6]. Coincident triggers between detectors are recorded by repeating this process with many time shifts. Subsequently, FAR is determined for detected signals based on the number of noise coincidences that exceed a certain statistical threshold within a given background time.

Ideally, one can utilize the time-shift technique to estimate the background and determine the statistical significance of GW signals in the hierarchical search method. However, this technique may not achieve optimal effectiveness owing to two main factors. First, there is the potential issue of a biased background when the second search is conducted solely on zero-lag (foreground) coincident triggers obtained from the first-stage search. This problem arises from insufficient background triggers, leading to a biased estimation of the GW event's FAR values detected in the second stage. Second, there is a tradeoff in computational advantages if all coincident triggers from the time-shift analysis of the first-stage search are followed for the second-stage search. While constructing a background by the time-shifting method, a larger number of triggers in the second stage may resolve the bias in FARs; it also increases the number of matched filtering operations, subsequently reducing the computational efficiency of the search.

Gadre et al. [25] addressed the issue of biased background estimation by proposing a heuristic approach that

involves assigning significance using a scaled coarse background. This method was later demonstrated in Soni *et al.* [31], showing that a scaled background could reasonably approximate the significance of GW events. This approach lacks a concrete foundational basis for background estimation.

In this work, we outline an approach for background estimation within a two-stage hierarchical search. Our methodology empirically models the noise distribution obtained in the second stage of the search, particularly after implementing the time-shifting method. The procedure goes as follows:

First, we conduct a coarse search to identify coincident triggers, encompassing foreground and background events as potential candidates for GW events. To mitigate excessive background triggers resulting from lower thresholds on single-detector statistics, we strategically reduce the number of time shifts in the first stage while maintaining a manageable increase in computational cost for matched filtering. The choice of time shift number (N_{shift}) is determined proportionally to the computational cost ratio between the two stages, resulting in a practical reduction without compromising the overall computational gain. For our analysis, we select $N_{\text{shift}} \gtrsim 148$ with a corresponding time shift of $T_{\text{shift}} \sim 5000$ s, applied to approximately 8.8 days of coincident data from LIGO's third observing run (O3). However, these values may be adaptable based on specific datasets. Our primary emphasis lies in identifying an optimal time shift value that yields a smooth background curve.

In the second stage, we delve into the matched filtering process within the nbhds of the previously identified candidates from the coarse search. Employing matched filtering with templates in these nbhds, we collect resultant triggers for further analysis, focusing on evaluating GW signal significance. This assessment involves constructing a background distribution through time-shifting the data at intervals of 5000 s. To enhance the accuracy of the noise distribution tail, we employ an empirical modeling technique. Specifically, we model the tail as a logarithmically decreasing curve with respect to ranking statistics.

We test the effectiveness of our background estimation approach with an extensive injection campaign. Our proposed background estimation method gives a comparable sensitivity with the two-detector PyCBC search for binaries in the low-chirp mass region while significantly reducing the computational costs associated with matched filtering operations. Notably, our pipeline demonstrates an impressive speed-up, performing nearly 13 times faster than the PyCBC search.

The paper is organized as follows: In Sec. II, we provide a comprehensive review of the search methodology employed by the flat search for a two-detector network. We review the use of matched filtering, delve into the data acquisition process and data-quality checks, explain the

derivation of single-detector and ranking statistics pertinent to the two-detector search, and briefly describe their approach to estimating the FAR. These definitions serve as the foundation for our subsequent discussions in Sec. III, where we review the hierarchical search method for a two-detector network. In Sec. IV, we present our new method for background estimation designed explicitly for the hierarchical search. In Sec. V, we examine the robustness of our background estimation approach by applying it to real data and comparing its consistency with the flat search. Furthermore, in Sec. V B, we compare the sensitivity of our search pipeline with that of the flat search. In Sec. V C, we evaluate the computational efficiency of our pipeline. Finally, we conclude our findings and draw inferences in Sec. VI.

II. REVIEW OF EXISTING TWO-DETECTOR FLAT SEARCH

The search for GW signals embedded in the noise $n(t)$ in a two-detector configuration is performed via a matched filtering technique using the PyCBC-toolkit [6]. This technique involves cross-correlating the output $s(t)$ of an interferometer with a modeled waveform known as template $h(t_c, \phi_c, \vec{\theta})$ in the frequency domain. The correlation is performed for different values of the coalescence time (t_c) and phase (ϕ_c), maximizing the matched-filter SNR $\rho(t)$ given the source parameters[1] ($\vec{\theta}$). Mathematically, this can be expressed as

$$\rho(t_c; \vec{\theta}) \equiv |(s, (1 + i)h(t_c, \phi_c = 0, \vec{\theta}))|. \tag{1}$$

In Eq. (1), (.,.) represents the scalar product between any two data time series $x(t)$ and $y(t)$ weighted by the one-sided power spectral density (PSD) $S_n(f)$ of the interferometer. The scalar product is defined as

$$(x, y) \coloneqq 4\mathcal{R}\left\{ \int_{f_{\text{low}}}^{f_{\text{high}}} \frac{\tilde{x}(f)\tilde{y}^*(f)}{S_n(f)} df \right\}. \tag{2}$$

Here, $\tilde{x}(f)$ and $\tilde{y}(f)$ denote the Fourier transforms of $x(t)$ and $y(t)$ respectively, and $\mathcal{R}$ represents the real part of the complex number. The integration is performed over a frequency range of f_{low} to f_{high}.

As the incoming signal parameters are not known in advance, matched-filter statistics are computed on data sampled at 2048 Hz using a flat template bank (see Table I). If the peak of the matched-filter SNR ρ is greater than or equal to a predefined threshold ρ_{thr} ($\rho_{thr} = 4$), a *trigger* is stored for further analysis. However, these triggers can be generated due to non-Gaussian transients in the data, resulting in many false alarms. Therefore, the triggers

---

[1]θ includes the component masses (m_1, m_2) and dimensionless spin vectors (s_{1z}, s_{2z}) of a spin-aligned binary system.

TABLE I. Summary of minimal match values and parameter ranges for coarse bank constructed using the hybrid geometric-random template placement algorithm [33,34] and the flat bank as adopted in the third gravitational wave transient catalog (GWTC-3) [2]. These banks are specifically designed to search for redshifted total mass within the range of $[2, 500]M_\odot$, encompassing dimensionless spin parameters for black holes ranging from -0.99 to 0.99 and for neutron stars ranging from -0.05 to 0.05. These parameter ranges have been selected to ensure consistency with the ranges employed by the PYCBC-broad search described in Abbott *et al.* [2].

Bank	Templates	Minimal match (MM)
Coarse	85,080	0.90
Flat	428,725	0.97

are down-weighted using their chi-square (χ_r^2) [35] and sine-Gaussian chi-square (χ_{sg}^2) values [36]. Using these chi-squares, single-detector statistics $\hat{\rho}$ are computed, given as

$$\hat{\rho} = \begin{cases} \tilde{\rho}(\chi_{sg}^2)^{-1/2} & \text{if } \chi_{sg}^2 \geq 6, \\ \tilde{\rho} & \text{otherwise,} \end{cases} \quad (3)$$

where $\tilde{\rho}$ is

$$\tilde{\rho} = \begin{cases} \frac{\rho}{[(1+(\chi_r^2)^3)/2]^{1/6}} & \text{if } \chi_r^2 \geq 1, \\ \rho & \text{otherwise.} \end{cases} \quad (4)$$

After reweighting, the surviving triggers are subjected to a coincidence test to identify instances of time coincidence with a high value of ranking statistics (Λ) [37] within the two-detector network. The statistic is given by

$$\Lambda = \frac{p(\vec{\kappa}|S)}{p(\vec{\kappa}|N)} \equiv \frac{p(\vec{\kappa}|S)}{r_{\vec{\theta}}^{HL} p(\vec{\theta}, \delta t_c, \delta\phi_c|N)}, \quad (5)$$

where Λ is the ratio of coincident event rate densities due to signal $p(\vec{\kappa}|S)$ and noise $p(\vec{\kappa}|N)$ for a coincident trigger's parameter denoted by $\vec{\kappa}$ that included includes the reweighted SNRs in the Hanford ($\hat{\rho}_H$), and Livingston ($\hat{\rho}_L$) detectors, the reduced chi-squared values ($\chi_{sg,H}^2, \chi_{sg,L}^2$), the time and phase differences of the coalescence ($\delta t, \delta\phi_c$) and the template parameter $\vec{\theta}$.

In the scenario of uncorrelated noise between the two detectors, the background noise events are expected to have a uniform distribution in the time and phase differences [37]. Consequently, $p(\vec{\theta}, \delta t_c, \delta\phi_c|N)$ can be treated as a constant. Therefore, in the allowed time window of coincidences τ_{HL} between two detectors, $p(\vec{\kappa}|N) \equiv r_{\vec{\theta}}^{HL}$ is the product of noise rate densities in each detector given by

$$r_{\vec{\theta}}^{HL} = 2\tau_{HL}\hat{r}_{\vec{\theta},H}(\hat{\rho}_H)\hat{r}_{\vec{\theta},L}(\hat{\rho}_L). \quad (6)$$

The signal rate density is obtained through Monte Carlo simulations as detailed in [7,37]. On the other hand, the individual noise rate densities are obtained by fitting the observed single-detector triggers to exponential functions above a fixed threshold of 6 on $\hat{\rho}$ and then calculating the fit parameters. To counteract the high variance in fit values due to the low number of triggers above the threshold, a moving average is applied to the fit parameters. Furthermore, the fit parameters are smoothed by averaging over nearby templates with similar values of effective spin, template duration, and symmetric mass ratio, following the method employed in Davies *et al.* [7].

Due to non-Gaussian features like glitches in the data, many loud coincidences may occur, posing a challenge in distinguishing true GW signals from noise and assessing their significance. Therefore, to measure the significance, FAR is computed by performing time shifts to the triggers by 0.1 s and recalculating the ranking statistics between the detectors. By generating a background of false coincidences (n_b) through multiple time shifts, the generated distribution is used to compute the FAR [6] as follows:

$$\text{FAR} = \frac{(1 + n_b(\Lambda_b \geq \Lambda_f^*))}{T_b}, \quad (7)$$

where T_b represents the observation time for the background estimation, Λ_b denotes the coincident statistic values of background triggers, and Λ_f^* for the foreground trigger that may or may not define the real GW event.

III. REVIEW OF TWO-STAGE HIERARCHICAL SEARCH

The hierarchical search involved a matched filtering search in two stages, as outlined in [31]. In the first stage, data segments sampled at 512 Hz are filtered using the templates from a coarse bank specified in Table I. Triggers with ρ and $\tilde{\rho}$ values exceeding 3.5 are collected for each detector in the network. For templates with a total mass greater than $30M_\odot$, the triggers are further re-weighted using χ_{sg}^2 to mitigate the impact of short-duration glitches on the noise. The selected triggers undergo coincidence testing where ranking statistics, denoted by Λ_1, are computed for foreground and background triggers.

Unlike in the flat search, a background is constructed in this stage by performing multiple time shifts with an interval of $T_{\text{shift}} = 5000$ s. This approach serves two primary purposes. First, it enables the examination of all coincident triggers, encompassing both foreground and background triggers, during the second stage search, where a finer search is conducted in the vicinity of the followed-up trigger's template parameter space. This comprehensive analysis predominantly accounts for noise coincidences in

the second-stage search, which were not previously accounted for in Soni *et al.* [31]. Second, conducting time shifts of $T_{\mathrm{shift}} = 5000$ s is more computationally efficient. This approach entails performing a finer search on fewer coincident triggers identified in the first stage, thereby avoiding a significant increase in the computational cost of matched filtering in the second stage.

The second stage search is conducted exclusively for those data segments with triggers having $\Lambda_1 \geq 7$. In this stage, these data segments are resampled at 2048 Hz and matched filtered against templates residing in the union of nbhds of their followed-up coarse templates. Ideally, nbhds for each followed-up coarse template could be generated on the fly. However, this process may increase the total computation time. Therefore, the nbhds for each coarse template are pre-generated as detailed in Sec. II (A2) of Soni *et al.* [31]. The number of templates within a nbhd typically ranges from approximately 10 to 100, as depicted by the discrete color scale in Fig. 1 on the chirp mass-effective spin plane. As seen from Fig. 1, the number of these templates is typically lower for low-mass binary systems. This happens because the boundaries of the coarse template's nbhds, defined by the minimal match ellipsoid, tend to extend beyond the actual parameter space. The effects of boundaries on the construction of nbhds are thoroughly discussed in [31].

In the second stage, the triggers with ρ and $\tilde{\rho}$ above 4 are collected from each detector in the network for further analysis. These triggers are again down-weighted with χ^2 and χ^2_{sg} like in the first stage search. The surviving triggers then undergo a coincidence test, where coincidence statistics Λ_2 is calculated. This statistic explicitly represents Λ associated with the second stage search. Note that when calculating $\hat{\rho}_\theta^{HL}$ in this stage, the fit parameters—smoothed over the effective spin, template duration, and symmetric mass ratio—tend to have large variances in their values due
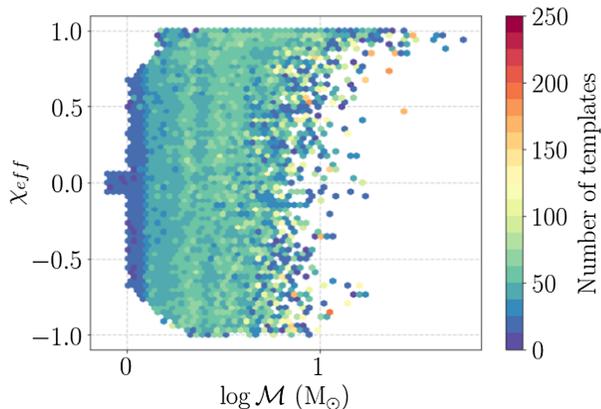
FIG. 1. Hexbin plot depicting the distribution of coarse templates in the logarithm of chirp mass ($\mathcal{M}$)—effective spin (χ_{eff}) plane. The color scale represents the total number of templates in the nbhd of each coarse template.

to an insufficient number of triggers. Therefore, we employ the strategy described in [31] of reusing the fit values from the "closest" coarse template to the corresponding trigger template.

The hierarchical search is expected to yield results in foreground and noise background candidates similar to the flat search. This expectation is primarily because both searches use the identical detector noise distribution, uniformly sampled at 2048 Hz. Even though the detector's underlying noise probability distribution is unknown, the statistical properties remain consistent in both searches. Additionally, both searches use banks characterized by consistent intrinsic source parameter ranges. Given the similar single-detector statistics, identical thresholds applied for the trigger selection, and coincident statistics, we at least expect the louder single-detector and coincident triggers to be consistent between the two searches. However, if the triggers from the foreground or background are faint, the hierarchical search might diverge from the flat search in its findings. The reason is that we only follow up coincident triggers from the first stage with $\Lambda_1 \geq 7$ in the second for a finer search, where we reidentify them with the best matching templates from their respective nbhds. Consequently, if triggers are faint in the first stage of the search, they may remain undetected due to our selection criteria.

Another scenario that may cause differences in the outputs of the two searches is when detected parameters, such as those of louder events or foreground or background triggers, vary due to the use of different templates. This is particularly relevant for templates from the high chirp mass region of the coarse bank. As shown in Fig. 1, templates within this high chirp mass region are distributed sparsely. Furthermore, by design, the number of templates in this region is fewer than in the flat bank. If a loud trigger generates from this region, a different coarse template parameter, and hence its nbhd, can give ranking statistics that may differ from those identified by the flat search. This difference is mainly due to the computation of Λ_2 in the second stage search, as elaborated in Sec. II C of Soni *et al.* [31]. Nevertheless, since the associated search cost in the high chirp mass range is low, a flat search can be implemented instead of a hierarchical search.

IV. A DIRECT METHOD FOR ESTIMATING THE FAR IN THE HIERARCHICAL SEARCH

The distribution of coincident triggers obtained in the second stage from time-shifted analysis is used for estimating FAR for GW events. To calculate FAR, we employ a hybrid approach that combines the time-shifted background with a fitting process applied to the tail of the distribution. This fitting process involves using a falling exponential model to account for true noise triggers. The application of the exponential-fitting procedure has previously been explored in estimating significance for single-detector

events, as demonstrated in Davies and Harry [38]. In the present study, we leverage this method to address the distribution of time-shifted coincidence triggers. The procedural approach is as follows:

First, we apply multiple time shifts with an interval of 5000 s to the triggers. This leads to the recalculation of the quantity Λ_2. As outlined in Capano *et al.* [39], obtaining an unbiased coincident noise distribution involves the removal of zero-lag coincident triggers from the background ones. Therefore, we adopt this approach to segregate the foreground coincidences from the time-shifted ones and obtain a distribution that exclusively contains the true noise triggers.

Applying time shifts with a 5000 s interval on approximately 8.8 days of data generates roughly one year of background time. In contrast, the background time generated by the flat search, using a time-shift interval of 0.1 s, spans a considerably longer duration. To ensure a fair comparison, we generate a comparable amount of background by extrapolating the tail of the cumulative distribution of Λ_2. This extrapolation allows us to estimate the background level for a time duration equivalent to the 0.1 s time-shift interval, enabling a meaningful assessment of the FAR estimation.

In an ideal scenario, the most systematic approach to testing the accuracy of our proposed method would involve verifying that the noise triggers do not have low FARs. This verification could be performed if a large dataset is analyzed or billions of time shifts are applied to a limited duration of data. However, these two approaches may not be computationally practical. Considering this, we have adopted a comparative strategy to assess the validity of our background estimation. Specifically, we compare the extrapolated noise background distribution obtained using the proposed method with the one obtained from a flat search.

Before proceeding with the extrapolation, we appropriately scale the background distribution by assuming that the number of noise coincidences in the second stage increases by a factor determined by the ratio of the time-shift intervals used in the hierarchical search (5000 s) and the flat search (0.1 s). This scaling ensures that the cumulative distribution of the hierarchical background matches that of the flat background, as shown in Fig. 2. Note that this scaling approach differs from that used in [31], where the hierarchical background was obtained by scaling the coarse-bank background to match the flat-bank background using sampling rate and template number argument. In the present work, the scaling is in total data volume, which is justifiable, as false alarm distribution is broadly independent of the length of the data.

Figure 2 also shows that the number of background triggers generated in hierarchical search is low compared to the ones obtained in the flat search. This discrepancy is mainly because of the stringent follow-up criteria applied during the first stage search and the limited number of time shifts implemented in the second stage search. Because of the former, fainter coincident triggers are likely to be missed in
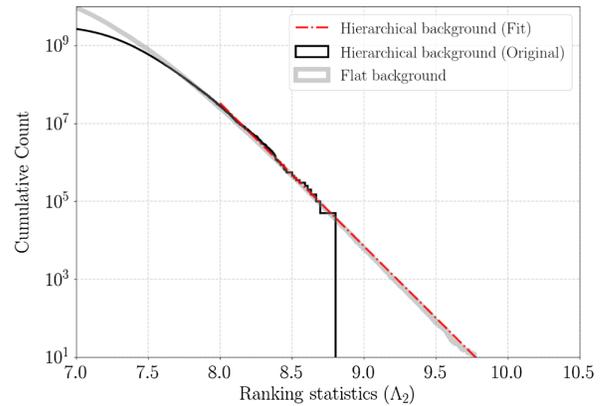
FIG. 2. Plot depicting the cumulative distribution of background events (excluding the foreground) versus ranking statistics for flat and hierarchical searches. The gray curve represents the background distribution obtained from a time-shift interval of 0.1 s in the flat search, while the black curve corresponds to a 5000 s time-shift interval in the hierarchical search. A linear fit to the hierarchical background (red) is superimposed to demonstrate that the cumulative number of events decreases linearly on a logarithmic scale with the ranking statistics.

the hierarchical search, which explains the divergence of the curve for $\Lambda_2 < 7.5$. Due to the latter, consequently, the maximum value attained by Λ_2 is approximately 8.8, which may not be sufficient for estimating the FAR. While the counts vary for low statistical values, the background appears consistent for higher statistics. Additionally, the background distribution obtained through the time-shifting method displays an approximately linear trend on a logarithmic scale with increasing Λ_2. This behavior suggests that linear regression methods could be used to extrapolate the curve and estimate trigger values for higher Λ_2 values.

Accordingly, we perform a linear regression on the cumulative distribution of the number of background coincident triggers represented on a logarithmic scale. For high $\Lambda_2 \gtrsim 8$, the linear regression can be written as

$$\log(n_{b,2}) = \alpha + \beta\Lambda_2, \tag{8}$$

where α and β are the fit parameters and $n_{b,2}$ is the number of background triggers in the second stage exceeding Λ_2. The linear regression assumes that the tail of the true noise coincidences follows an exponential dependence, consequently implying that $\beta < 0$. This assumption remains valid even when more coincident noise triggers are followed up from the first stage. Such a scenario may happen if more time shifts are performed instead of performing a few hundred corresponding to a time interval of 5000 s. However, this approach may also increase the number of matched filtering operations in the second stage and reduce the computational advantages of hierarchical search over flat search. A straightforward calculation, detailed in the Appendix, substantiates the anticipated presence of this

feature. A straightforward calculation, detailed in the Appendix, substantiates the anticipated presence of this feature.

Note that the exponential fit method may not be suitable for cases where loud background triggers, which are not eliminated from the data, increase the background. This problem has also been observed in generic flat searches and primarily occurs when very short-duration templates produce triggers in the presence of glitches, resulting in an inflated background. Such instances should be carefully examined before proceeding with the fitting procedure. To mitigate such problems in hierarchical search, one approach could involve removing these loud background triggers from the coarse background. Since only coincident triggers are followed up in the second stage, the probability of loud noise coincidences affecting the second-stage background is relatively low. Once these triggers are removed and the second-stage search is performed, the fitting procedure can be applied.

Once we have obtained the fitted noise distribution, we compute FAR for the foreground events detected in the second stage of the search. This involves utilizing Eq. (7) and the fitted distribution to determine the number of background coincidences that have a Λ_2 value greater than or equal to that of the statistic value foreground trigger.

V. APPLICATION TO REAL DATA

We now employ a hierarchical search pipeline on real data, as elaborated in Sec. III. We apply the methodology outlined in Sec. IV to calculate FARs for the detected events in our analysis. For comparison, we will utilize a two-detector PyCBC-broad search pipeline.

For this study, we analyzed approximately 8.8 days of data extracted from May 12 to May 21, 2019, during the third observational run of the twin LIGO–LHO-LIGO Hanford Observatory and LLO-LIGO Livingston Observatory. We used the strain data obtained from GWDataFind [40] for both the detectors. This data encompasses a mixture of Gaussian and non-Gaussian noise features called *glitches*. The presence of the latter, often attributed to diverse instrumental artifacts, requires their removal before the matched-filtering step. Therefore, the timestamps displaying suboptimal quality or instances of data unavailability were removed using Category 1 and Category 2 vetoes [41]. Moreover, transient short-duration artifacts in the data were eliminated using a gating method [6]. Although most known noise artifacts can be removed before the matched filtering step, some glitches persist. These were subsequently filtered out through the application of different chi-squares criteria as described in Sec. II.

Once the timestamps unsuitable for astrophysical searches were removed, we ran both search pipelines to initiate the analysis. For the flat search, we employed the flat bank previously constructed for the PyCBC-broad search described in Abbott *et al.* [2]. For hierarchical search, we

used pre-computed coarse and nbhd banks as elaborated upon in Soni *et al.* [31]. These sets of banks were specifically designed to explore the parameter spaces of the BBH, BNS, and NSBH systems. Further, the banks were tailored to cover a wide range of total masses, from 2 to $500 M_\odot$, as well as component spins within the respective ranges of 0 to 0.9 for black holes and 0 to 0.05 for neutron stars. Spin precession was not considered in this study.

A. Results

The analysis resulted in the identification of numerous foreground and background coincidences. We assessed the significance of foreground triggers using the method outlined in Sec. IV. As depicted in Fig. 3, the FARs of the background triggers obtained from the background fit closely match those derived from the traditional time-shifting method in the hierarchical search up to a threshold of approximately $\Lambda_2 \sim 8.8$. Furthermore, the fit aligns well with the curve obtained from the flat search, wherein a larger number of time shifts, equivalent to an interval of 0.1 s, are employed. This outcome demonstrates the effectiveness of our proposed approach for background estimation, as elaborated in Sec. IV.

We also examined the significance of foreground events identified in our analysis. Table II summarizes our findings from the two searches. Our method identified all events characterized by high network SNRs with consistent FAR values within a factor of a few compared to those obtained from the flat search. We were able to recover two previously detected GW events—GW190521_074359 and GW190519_153544, as cataloged in Abbott *et al.* [2]. Additionally, we identified two other low-significance
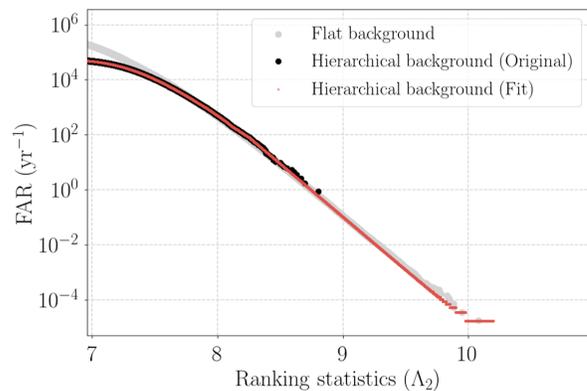
FIG. 3. A comparison plot of the FAR as a function of ranking statistics of the background triggers obtained in flat and hierarchical search methods. The FAR distribution with and without exponential fit to the hierarchical background distribution is shown in black and red, respectively. The FAR distribution for background from the flat search is overlaid (gray) to demonstrate that the linear-fit approximation adequately matches the FAR values between the two search methods.

TABLE II. Results from a two-detector analysis over data duration from May 12 to May 21, 2019, using flat and hierarchical searches. Listed events have a FAR less than 1 per year and are ordered in descending order based on their network SNR ($\hat{\rho}_T = \sqrt{\rho_H^2 + \rho_L^2}$). The table also provides chirp mass ($\mathcal{M}$) for each identified event. The FARs of the detected events in the flat search were determined using the time-shift method, whereas those for the hierarchical search were determined by our proposed method described in Sec. IV. The redetection of these events is consistent with the GWTC-3 [2].

Event	Flat			Hierarchical		
	FAR (yr^{-1})	$\hat{\rho}_T$	$\mathcal{M}$ ($M_\odot$)	FAR (yr^{-1})	$\hat{\rho}_T$	$\mathcal{M}$ ($M_\odot$)
GW190521_074359	$<1.76 \times 10^{-5}$	24.03	30.56	$<1.76 \times 10^{-5}$	23.31	24.32
GW190519_153544	$<1.76 \times 10^{-5}$	13.19	73.04	$<3.53 \times 10^{-5}$	12.55	71.25
GW190513_205428	3.7×10^{-4}	11.61	32.52	1.9×10^{-4}	11.79	28.89
GW190517_055101	9.0×10^{-3}	10.13	44.15	3.4×10^{-2}	10.45	36.74

events, GW190513_205428 and GW190517_055101, also reported in the GWTC-3.

As seen in Table II, the four events have varying network SNR values in both searches. These variations are attributed to the differences in their recovered coincident templates, which yield distinct chirp mass values. As illustrated in Fig. 1, the number of coarse templates in the coarse bank in the high chirp mass range is less. As a result, the chances of precisely identifying the template in the nbhd of a coarse template from high-chirp mass regions could be lower, leading to potential differences in the recovered parameters.

These findings are also illustrated in Fig. 4, where we present a comparison of the inverse false alarm rates (IFARs) for the recovered events in both hierarchical and flat search methodologies. This comparison suggests that the hierarchical search yields a candidate list similar to the flat search, though there are some differences for very quiet events. Additionally, our method seems to provide similar IFAR estimates for notable events in the O3 data.
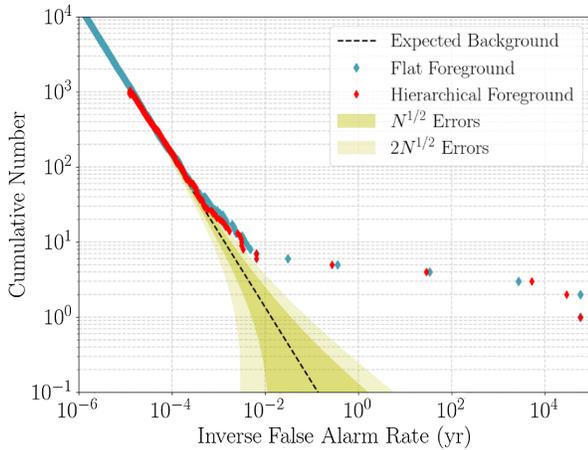
FIG. 4. Comparison of the cumulative histogram of the inverse false alarm rate (1/FAR) for the foreground triggers obtained in the flat (blue) and hierarchical (red) searches. The dashed line represents the expected background for a given observation time T, while the yellow-shaded regions represent Poisson errors estimated for the flat search.

The data quality employed in the analysis could affect the fitting procedure for acquiring the background. To assess the efficacy of our methodology across various instances of noise, we applied it to multiple datasets collected during the first two observing runs of the twin LIGO. Specifically, our analysis concentrated on approximately 5.3 days of concurrent data from O1 and 5.5 days from O2. Our results, as shown in Fig. 5, suggest a consistent alignment between the background trigger rates from our method and those from the flat search, indicating the reliability of our background estimation approach.

B. Sensitivity comparison

We now compare the sensitivities of two pipelines through an extensive injection campaign. We use sensitive volume-time (VT) [42] as a metric to compare the sensitivities of the search pipeline for a simulated population of binary injections in the data.

The VT product for a search pipeline measures the projected number of detectable signals originating from a population of binary systems that exceeds a predefined level of statistical significance. Assuming the binary merger rate remains constant, the ratio of the VT product between the two search pipelines provides insight into their relative sensitivities. Typically, this evaluation is carried out using Monte Carlo simulations [42].

For this study, we systematically injected various signals corresponding to the BBH, BNS, and NSBH mergers. We then observed their recovery using both flat and hierarchical searches, as described in Secs. II and III, respectively. Specifically, we targeted sources with intrinsic parameters as listed in Table III. The parameter ranges selected for these injections were inspired by the PyCBC analysis conducted during O3.

To simulate these sources, we generated non-precessing quasi-circular and quadrupolar GW signals using the SpinTaylorT4 [43] waveform model for BNS sources and the SEOBNRv4_opt [44] waveform model for BBH and NSBH sources. These signals were injected at a time step interval of 100 s within the data. For simplicity,
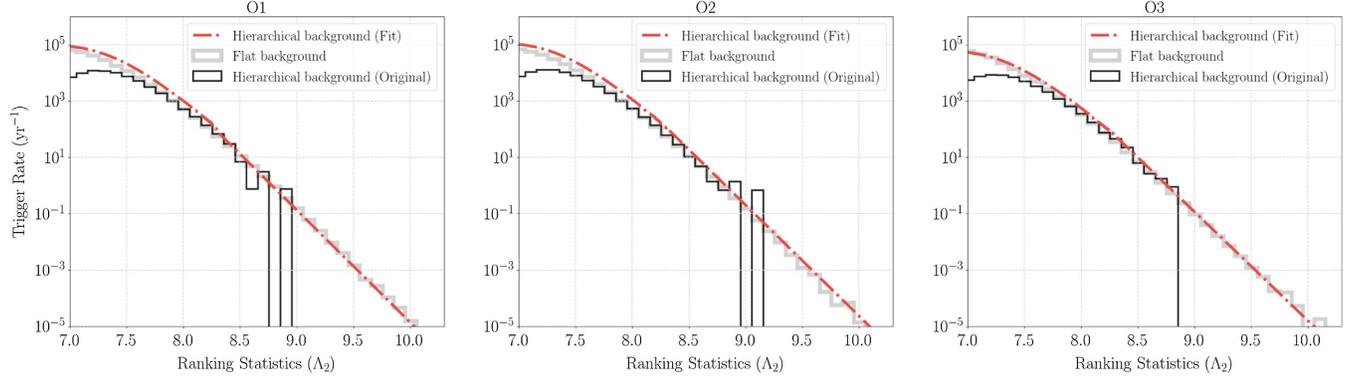
FIG. 5. Comparison of the histograms of background trigger rates (excluding the foreground) versus ranking statistics for O1, O2, and O3 data. The black curve represents the true background obtained from time-shifted triggers across the detectors. The red curve represents the background obtained after fitting the true background trigger distribution. The background trigger rate from a flat search is overlaid in gray for comparison.

we distributed these signals isotropically all over the sky and chirp distances ranging from 5 to 400 Mpc.

We observed the recovered injections and used VT product computations to examine the sensitivities of the pipelines. Our findings, illustrated in Fig. 6, suggest that the VT ratios between the two pipelines for low chirp masses and across all the IFAR bins are close to unity. Specifically, the values fluctuate between 0.99 ± 0.144 and 1.02 ± 0.141 for chirp masses ~ 1.4–$20 M_\odot$ at an IFAR of 10 years. This result indicates that the sensitivity of the hierarchical search is comparable to that of the flat search, particularly for binary systems characterized by low chirp masses. However, the sensitivity of the hierarchical search begins to diminish for chirp masses greater than $20 M_\odot$. This reduction is primarily attributed to the lower density of coarse templates within the higher chirp mass range and the domain of low effective spins, as shown in Fig. 1.

C. Computation efficiency

We demonstrated that our background computation has good signal detection capabilities while maintaining a

TABLE III. The table presents the parameter distributions and corresponding ranges for m_1, m_2 (redshifted component masses), and χ_1 and χ_2 (dimensionless effective spins) for each source category. The specified parameter intervals align with those used in Abbott *et al.* [2].

Source	Parameter	Distribution	Range
BBH	m_1, m_2	Log-uniform	2.5–150$M_\odot$
	χ_1, χ_2	Uniform	0–0.9
BNS	m_1, m_2	Uniform	1–2.5$M_\odot$
	χ_1, χ_2	Uniform	0–0.4
NSBH	m_1	Log-uniform	2.5–97.5$M_\odot$
	m_2	Log-uniform	1–2.5$M_\odot$
	χ_1	Uniform	0–0.9
	χ_2	Uniform	0–0.4

search sensitivity comparable to the flat search in low-chirp mass regions. To further evaluate the performance of our search pipeline, we now turn our attention to its computational efficiency.

The computational efficiency of a search pipeline relies heavily on the speed at which matched filtering operations, which inherently involve performing fast Fourier transforms (FFTs), are executed. This aspect can be evaluated by directly utilizing the Cooley-Tukey algorithm [45] to compute the number of floating-point operations involved.

Let us assume that the data segment is sampled at a frequency of f Hz and has a t_{seg} s duration. In this case, the number of FFT operations can be expressed as $\kappa N \log_2(N)$, where $N = f t_{\text{seg}}$ and κ is approximately a few. This operation is repeated for several templates denoted by N_{temp}, resulting in an actual computational cost of approximately $\kappa N_{\text{temp}} N \log_2(N)$. It means that the computational cost of matched filtering scales with the data sampling rate and number of templates.

Ideally, the speed-up factor in the matched filtering computation between the flat and hierarchical search methods can be estimated by calculating the ratio of the total cost in the flat search to the combined total cost in both stages of the hierarchical search. This speed-up can be represented as

$$\text{speed-up} \approx \frac{N_{\text{seg}} N_{\text{temp}}^{\text{flat}} O_{\text{flat}}}{N_{\text{seg}} N_{\text{temp}}^{\text{stage1}} O_{\text{coarse}} + N_{\text{temp}}^{\text{stage2}} O_{\text{fine}}}, \quad (9)$$

where,

$$O_{\text{flat}} = \kappa f_{\text{flat}} t_{\text{seg}} \log(f_{\text{flat}} t_{\text{seg}}),$$
$$O_{\text{coarse}} = \kappa f_{\text{coarse}} t_{\text{seg}} \log(f_{\text{coarse}} t_{\text{seg}}),$$
$$O_{\text{fine}} = \gamma O_{\text{flat}}.$$
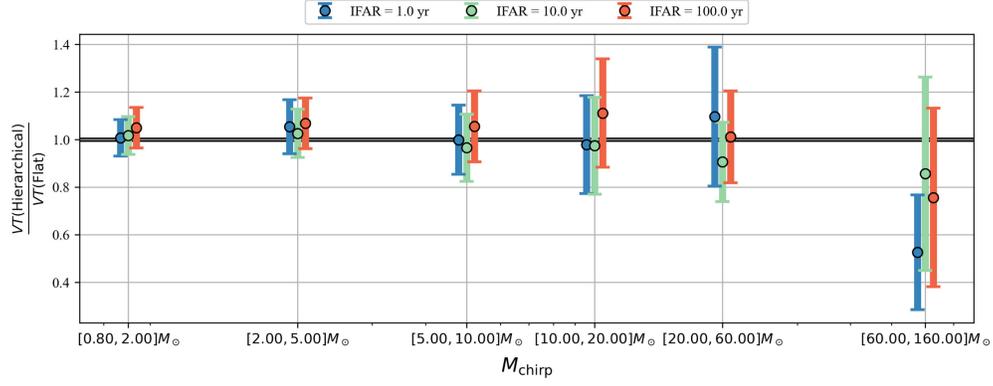
FIG. 6. Plot depicting volume-time (VT) ratio between hierarchical and flat search.

Here, N_{seg} represents the total number of overlapping segments analyzed. The data sampling rates are denoted by f_{flat} (2048 Hz) for the flat search and f_{coarse} (512 Hz) for the first stage of the hierarchical search. While the matched filtering cost dominates the analysis of a data segment with a large template bank, overhead costs are associated with each segment, for example, for loading and initial data processing, which are independent of the number of templates. The overhead cost per template can become comparable to or even greater than the matched filtering cost when the number of templates is small. In the above formula, factor γ accounts for this.

We previously discussed a comparison of the matched filtering computational costs between the hierarchical and flat search pipelines using O3 data from the Hanford and Livingston detectors. In our analysis, the data from the Hanford detector was split into 176 blocks and that from the Livingston detector into 197 blocks. Each block was then segmented into overlapping durations of 512 s each.

In the flat search, each segment of the data block sampled at a rate of f_{flat} was filtered with $N_{\text{temp}}^{\text{flat}} = 428{,}725$ templates. However, the number of templates and data sampling rates varied between the two stages of the hierarchical search. In the first stage, the data segments were sampled at f_{coarse} and filtered with 85,080 templates denoted by $N_{\text{temp}}^{\text{stage1}}$. During the second stage of the search, despite the data segments being sampled at f_{flat} rates, the number of templates varied across segments, leading to fewer FFT operations than in the flat search. This variability in the number of templates per segment resulted from the number of templates in the union of the nbhds that were followed. Consequently, the number of templates fluctuates for each data block, as shown in Fig. 7.

If the number of templates used by each segment were large, γ would be close to 1. However, we find that for a typical data block, for the Hanford (Livingston) detector, the CPU time taken for flat search, with $428{,}725 \times 10$ (segments in the block) $\sim$4 million application of templates to data, was 66.52 hr (73.15 hr). In contrast, for the same block, in the second stage, with a total of 46,369

applications of templates to data, the CPU time taken was 1.88 hr (1.72 hr). If we compare the computation time per segment per template, one can see that the flat search ($\sim$0.05 s per template per segment) was approximately three times less expensive than the nbhd search ($\sim$0.15 s per template per segment), that is, $\gamma \sim 3$. We checked for
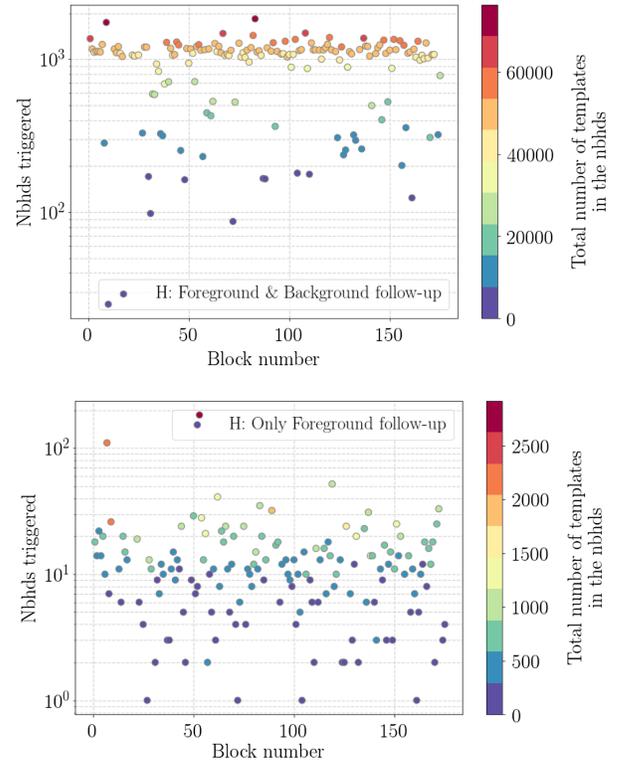
FIG. 7. Plot depicts the number of nbhds triggered per block while performing the second-stage search for data from Hanford (H). The color bar represents the total number of templates in the nbhds triggered for a block. The number of nbhds triggered and the corresponding number of templates in the nbhds is larger when foreground and background triggers are followed up in the second stage (top panel). However, these numbers are lower when only the second search is performed on the foreground triggers alone (bottom panel).

TABLE IV. Table depicting the CPU core hour for the matched filtering jobs in flat and hierarchical searches. The numbers are estimated separately for the two detectors–Hanford (H) and Livingston (L) in the analysis.

	CPU core hour	
Search pipeline	Hanford	Livingston
Flat	10,685.11	12,289.90
Hierarchical with background follow-up	712.89	790.71
Hierarchical with only foreground follow-up	452.96	529.21

different data blocks and γ remained nearly the same for similar data block lengths.

In our study, with t_{seg} set at 512 s and N_{seg} set at 1493 for the Hanford detector (or 1731 for Livingston), the total number of templates in the second stage of the hierarchical search was approximately 6,274,846 (or 6,235,197 for Livingston). The resulting speed-up in the matched-filtering computation was approximately 18.36 (or 18.38 for Livingston). Accounting for the overhead computation time per segment with $\gamma \sim 3$, the estimated speed-up, 13.58 for Hanford and 14.07 for Livingston, matches the actual CPU obtained by comparing the real CPU times.

We also compared the computational efficiency of our search pipeline using a commonly employed metric in high-performance computing environments, that is, CPU core hours. We effectively distributed matched filtering operations across multiple CPU cores by leveraging the parallelization features of the PyCBC toolkit. We found that the CPU core hour requirements for hierarchical search were consistently 14–15 times lower than those for flat search for both detectors, as shown in Table IV. This indicates a considerable reduction in computational resources and processing time associated with hierarchical searches.

VI. CONCLUSION AND DISCUSSION

Since its inception, the hierarchical search strategy has aimed to improve the efficiency of matched filtering by performing multi-stage searches using multiple banks of varying densities. Although it offers several advantages, especially in conducting computationally expensive searches for compact binaries such as sub-solar masses, accurately estimating an unbiased background has proven to be a persistent challenge. In particular, there is a constant tug-of-war between optimizing the computational cost of matched filtering and obtaining a proper noise background for assigning significance to detected candidates.

In this work, we have attempted to address this challenge by introducing a minor computational tradeoff. In Sec. IV, we outline a method to estimate an unbiased noise background. This method uses a hybrid approach that combines the time-shifted background distribution and empirically models its tail based on the assumption that the distribution contains only noise coincidences. The effectiveness of our scheme was thoroughly demonstrated in Sec. V, where we found that our estimated FARs of true events from O3 closely align with those obtained through flat search. This result indicates the reliability and accuracy of our method for capturing the true significance of GW events. Because the tail of the distribution was obtained by extrapolation, one may argue that the background estimate in this region may contain a certain level of inaccuracy. However, the tail corresponds to loud detections, and any minor discrepancies in the FAR for these occurrences are likely to hold little significance when making inferences about astrophysical implications. Note that the background for the lower-ranking statistic ($\lesssim 8.5$), where distinguishing signals from noise presents a challenge, our estimations are derived directly from the time-slides, without involving any extrapolation, and hence, should be as reliable as the flat search. If necessary, this lower-ranking statistic region can be expanded further, perhaps up to a ranking statistic of approximately ~ 9. This expansion would involve increasing the number of time shifts, incurring additional computation time. However, despite this increase in computational demand, the overall computational efficiency of the entire search is still significantly advantageous.

The significance of employing hierarchical search strategies for low-chirp mass sources becomes even more apparent based on our VT comparison results, as presented in Sec. V B. We observed that with the current detection statistics and template bank configurations, the sensitivity of detecting lower-mass binaries is similar to that of flat search. However, in the case of higher chirp mass ranges, the sensitivity is reduced owing to the sparsity of templates in both the coarse bank and nbhd bank. Consequently, precise measurements of masses and spins are challenging for very short-duration signals with a hierarchical search. Thus, it is advisable to consider a more targeted search approach for such parameter ranges. Because the total number of templates in higher masses is much smaller than that for lower masses, we recommend performing the standard flat search for high masses, which adds an insignificant amount to the total computation cost. Incidentally, dedicated searches focused on high-mass binaries are performed routinely [2]. Therefore, no additional effort may be necessary to improve computational efficiency in the above context for high-mass binaries.

Although our approach has a minor reduction in computational gains compared to the previous implementation [31], where the background is estimated by scaling, this tradeoff is more than justified. The current pipeline, while being reliable, manages to substantially reduce the computational cost of matched filtering by an order of magnitude, and we find a computational speed-up of ~ 13 for O3

data. This ensures that the proposed method is efficient and practical for real-world GW detection.

It is important to note that although our proposed method is effective for cases where multiple detectors detect GW signals, it may require careful attention and consideration for situations where the signal is captured by only one detector. In such cases, our background estimation may be susceptible to non-Gaussian artifacts caused by signal contamination. To address this potential issue, it is crucial to carefully extract and account for contamination, thus ensuring the reliability of our method, even in such challenging scenarios.

In conclusion, our study introduces a robust approach for obtaining an unbiased background in a hierarchical search, a powerful and efficient tool for GW search. By effectively balancing computational efficiency and accurate background estimation, our method paves the way for new avenues of research in the study of compact binaries.

APPENDIX: ANALYTICAL CALCULATION FOR COINCIDENT NOISE DISTRIBUTION IN GAUSSIAN NOISE

In this section, we attempt to understand the coincident noise distribution in the case of stationary Gaussian noise.

The analytical calculation presented in this section substantiates the anticipated presence of linearly falling background distribution that we discussed in Sec. IV. This exercise aims to gain insights into the statistical distribution of detection statistics, particularly the quadrature sum of matched filter SNRs, in the context of Gaussian-colored noise.

Suppose we aim to detect a GW signal embedded in stationary Gaussian noise, characterized by N independent random variables. To accomplish this, we perform matched filtering on a data segment with a length of 512 s, sampled at 2048 Hz, using a template that consists of two orthonormal polarizations, h_o and $h_{\pi/2}$. For the sake of simplicity, we assume that the occurrence of these signals is extremely rare within the data. The output of the matched filter can be regarded as the projection of the data vector onto the templates, as given by

$$\rho = \sqrt{c_o^2 + c_{\pi/2}^2}, \tag{A1}$$

where ρ represents the SNR obtained from the matched filter, while c_o and $c_{\pi/2}$ denote the projection coefficients onto h_o and $h_{\pi/2}$, respectively. Since the two polarizations are orthonormal, c_o and $c_{\pi/2}$ are normally distributed. The matched filter generates N' samples of ρ over time, and these samples follow a Rayleigh distribution, as shown in Fig. 8.

The samples produced by the matched filtering process are typically weakly correlated. Due to the weak correlations, we can assume that the cyclic operation of matched filtering generates $N < N'$ statistically independent Rayleigh variables. In such cases, the probability distribution of the maximum of N independent Rayleigh variables is given by

$$p_N(\rho) = N\rho e^{-\rho^2/2}(1 - e^{-\rho^2/2})^{N-1}. \tag{A2}$$

The above pdf can be easily obtained by taking the product of N Rayleigh distribution functions and then differentiating the product. For a large value of N, Eq. (A2) can be approximated to a simpler form amenable to easy analytical manipulations as

$$p(\rho) \simeq N\rho e^{-(\rho^2/2 + Ne^{-\rho^2/2})}. \tag{A3}$$

Let y_1 and y_2 represent two Rayleigh deviates corresponding to the SNRs from Hanford and Livingston, respectively. The probability distribution functions for y_1 and y_2 based on Eq. (A3) are given by

$$p(y_1) = N_1 y_1 e^{-(y_1^2/2 + N_1 e^{-y_1^2/2})} \tag{A4}$$

$$p(y_2) = N_2 y_2 e^{-(y_2^2/2 + N_2 e^{-y_2^2/2})}, \tag{A5}$$
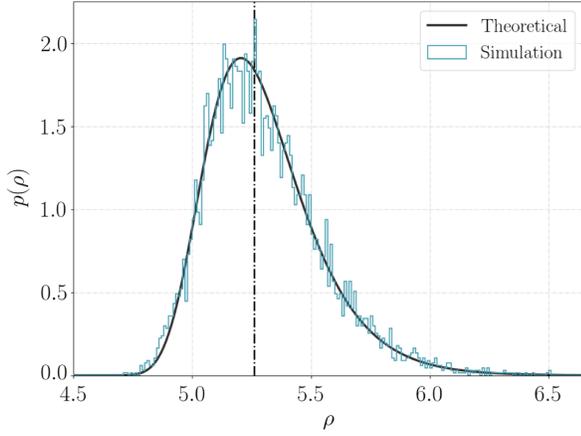
FIG. 8. Probability density function of ρ represented as a histogram (in blue). An ideal Rayleigh distribution is overlaid in black, providing a visual comparison. The vertical dotted line represents $\rho = 5.3$.

where N_1 and N_2 represent the maximum number of independent Rayleigh variables in each detector.

Let us define the network SNR as the coincident statistics in this case. Therefore, if we set $N = N_1 = N_2$, then the joint probability distribution function, denoted as $g(y_1, y_2)$, can be expressed as the product of the individual PDFs for y_1 and y_2 as

$$g(y_1, y_2) = p(y_1)p(y_2) \tag{A6}$$

$$= \frac{N^2}{2} y_1 y_2 e^{-\{\frac{1}{2}(y_1^2 + y_2^2) + N(e^{-y_1^2/2} + e^{-y_2^2/2})\}}. \tag{A7}$$

Consider solving the above problem in polar coordinates. That means if $y_1 = \rho_c \cos\phi$ and $y_2 = \rho_c \sin\phi$, $g(y_1, y_2)$ (i.e., $\rho_c = \sqrt{y_1^2 + y_2^2}$) will become

$$g(\rho_c, \phi) = \frac{N^2}{2} \rho_c^2 \sin 2\phi \, e^{-\rho_c^2/2}$$
$$\times \left[e^{-N\{e^{-\frac{\rho_c^2}{4}(1+\cos 2\phi)} + e^{-\frac{\rho_c^2}{4}(1-\cos 2\phi)}\}} \right]. \tag{A8}$$

To obtain the joint PDF as a function of ρ_c, we marginalize Eq. (A8) over ϕ, that means

$$g(\rho_c) = \frac{N^2}{2} \rho_c^3 e^{-\rho_c^2/2}$$
$$\times \int_0^{\pi/2} d\phi \, \sin 2\phi \left[e^{-N_c\{e^{-\frac{\rho_c^2}{4}(1+\cos 2\phi)} + e^{-\frac{\rho_c^2}{4}(1-\cos 2\phi)}\}} \right]. \tag{A9}$$

Setting $\cos 2\phi = u$

$$g(\rho_c) = \frac{N^2}{4} \rho_c^3 e^{-\rho_c^2/2} \int_{-1}^{1} du \, e^{-N\{e^{-\frac{\rho_c^2}{4}(1+u)} + e^{-\frac{\rho_c^2}{4}(1-u)}\}}$$
$$= \frac{N^2}{2} \rho_c^3 e^{-\rho_c^2/2} \int_0^1 du \, e^{-N\{e^{-\frac{\rho_c^2}{4}(1+u)} + e^{-\frac{\rho_c^2}{4}(1-u)}\}}. \tag{A10}$$
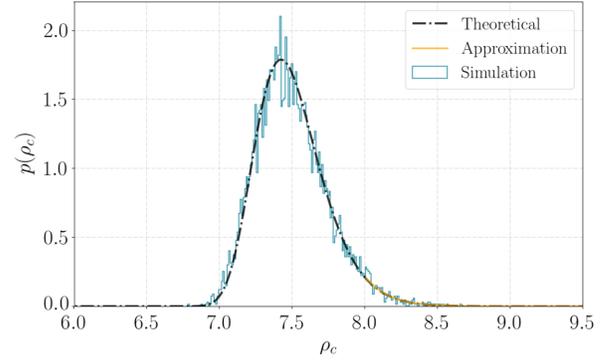
FIG. 9. Probability density function for network SNR as a histogram obtained from simulations (blue) and by numerically integrating Eq. (A10) (black).

It can be shown that the peak of this distribution occurs at $\sim\sqrt{4\log N}$. Although the above integration can be numerically performed, we derive a semianalytical fit for the tail of the distribution at large values of $\rho_c \gg \sqrt{4\log N}$, where it is observed that the term $f(u) = e^{-N\frac{\rho_c^2}{4}(1-u)}$ dominates the integral. The function $f(u)$ is approximately shaped like a trapezium: for low values of u, $f(u) \simeq 1$ and $\to 0$ as $u \to 1$. We write the integral as J:

$$J = \int_0^1 du \, f(u). \tag{A11}$$
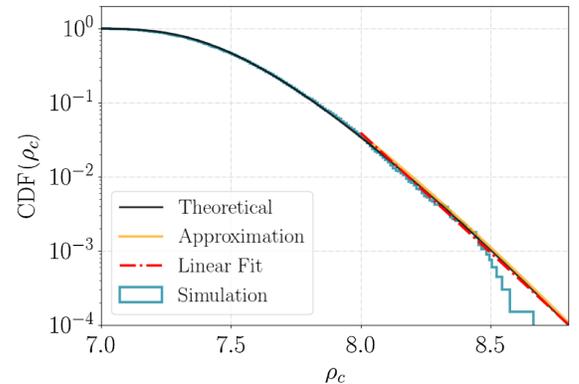
FIG. 10. Cumulative density distributions obtained by integrating Eqs. (A10) (black) and (A12) (orange), along with a histogram (blue) from simulated results. A linear fit (red) is overlaid in the logarithmic scale for comparison.

We approximate $f(u)$ using the trapezoidal rule. In this approximation, we set $f(u_o) = \frac{1}{2}$ at the mid-height of the trapezium, which has a unit height. Solving for u_o yields the integral J

$$u_o = 1 - \frac{4}{\rho_c^2}(\log N - \log \log 2) \equiv J. \qquad (A12)$$

From Fig. 9, it is evident that the approximation derived from Eq. (A12) aligns well with Eq. (A10) and the network SNR distribution obtained from the simulation. This

observation prompts us to examine the cumulative distribution, as depicted in Fig. 10.

Notably, the tail of the cumulative density function displayed in Fig. 10, derived from Eqs. (A10) and (A12), exhibits a characteristic pattern that can be approximated by linearly decreasing values of ranking statistics when plotted on a logarithmic scale. This finding forms a benchmark for effectively modeling background distribution acquired through time-shifting triggers in real analysis.

This exercise shows that the tail of the distribution in the general case has similar features as in this tractable case of Gaussian noise.

[1] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116,** 061102 (2016).

[2] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Phys. Rev. X **13,** 041039 (2023).

[3] B. Abbott *et al.*, Phys. Rev. Lett. **119,** 161101 (2017).

[4] B. P. Abbott *et al.*, Astrophys. J. Lett. **892,** L3 (2020).

[5] R. Abbott *et al.*, Astrophys. J. Lett. **915,** L5 (2021).

[6] S. A. Usman, A. H. Nitz, I. W. Harry, C. M. Biwer, D. A. Brown, M. Cabero, C. D. Capano, T. D. Canton, T. Dent, S. Fairhurst *et al.*, Classical Quantum Gravity **33,** 215004 (2016).

[7] G. S. Davies, T. Dent, M. Tápai, I. Harry, C. McIsaac, and A. H. Nitz, Phys. Rev. D **102,** 022004 (2020).

[8] T. Adams, D. Buskulic, V. Germain, G. M. Guidi, F. Marion, M. Montani, B. Mours, F. Piergiovanni, and G. Wang, Classical Quantum Gravity **33,** 175012 (2016).

[9] F. Aubin, F. Brighenti, R. Chierici, D. Estevez, G. Greco, G. M. Guidi, J. V., F. Marion, B. Mours, E. Nitoglia *et al.*, Classical Quantum Gravity **38,** 095004 (2021).

[10] C. Messick, K. Blackburn, P. Brady, P. Brockill, K. Cannon, R. Cariou, S. Caudill, S. J. Chamberlin, J. D. E. Creighton, R. Everett *et al.*, Phys. Rev. D **95,** 042001 (2017).

[11] S. Sachdev, S. Caudill, H. Fong, K. L. R. Lo, C. Messick, D. Mukherjee, R. Magee, L. Tsukad *et al.*, arXiv:1901.08580.

[12] C. Hanna, S. Caudill, C. Messick, A. Reza, S. Sachdev, L. Tsukada, K. Cannon, K. Blackburn, J. D. E. Creighton, H. Fong, P. Godwin, S. Kapadia, T. G. F. Li, R. Magee, D. Meacher, D. Mukherjee, A. Pace, S. Privitera, R. K. L. Lo, and L. Wade, Phys. Rev. D **101,** 022003 (2020).

[13] Q. Chu, Ph.D. thesis, The University of Western Australia, 2017.

[14] B. S. Sathyaprakash and S. V. Dhurandhar, Phys. Rev. D **44,** 3819 (1991).

[15] S. V. Dhurandhar and B. S. Sathyaprakash, Phys. Rev. D **49,** 1707 (1994).

[16] S. V. Dhurandhar and B. F. Schutz, Phys. Rev. D **50,** 2390 (1994).

[17] B. J. Owen and B. S. Sathyaprakash, Phys. Rev. D **60,** 022002 (1999).

[18] J. Aasi, B. P. Abbott, R. Abbott, T. Abbott, M. R. Abernathy, K. Ackley, C. Adams, T. Adams, P. Addesso *et al.*, Classical Quantum Gravity **32,** 074001 (2015).

[19] F. Acernese, M. Agathos, K. Agatsuma, D. Aisa, A. Allemandou, N. Allocca, J. Amarni, P. Astone, G. Balestri *et al.*, Classical Quantum Gravity **32,** 024001 (2014).

[20] T. Akutsu, M. Ando *et al.*, Nat. Astron. **3,** 35 (2019).

[21] B. Iyer *et al.*, LIGO-India Technical Report, 2011, https://dcc.ligo.org/LIGO-M1100296/public.

[22] M. Saleem, J. Rana, V. Gayathri, A. Vijaykumar, S. Goyal, S. Sachdev, J. Suresh, S. Sudhagar, A. Mukherjee, G. Gaur, B. Sathyaprakash, A. Pai, R. X. Adhikari, P. Ajith, and S. Bose, Classical Quantum Gravity **39,** 025004 (2021).

[23] B. P. Abbott, R. Abbott *et al.*, Classical Quantum Gravity **34,** 044001 (2017).

[24] I. Harry, S. Privitera, A. Bohé, and A. Buonanno, Phys. Rev. D **94,** 024012 (2016).

[25] B. Gadre, S. Mitra, and S. Dhurandhar, Phys. Rev. D **99,** 124035 (2019).

[26] R. Dhurkunde, H. Fehrmann, and A. H. Nitz, Phys. Rev. D **105,** 103001 (2022).

[27] S. D. Mohanty and S. V. Dhurandhar, Phys. Rev. D **54,** 7108 (1996).

[28] S. D. Mohanty, Phys. Rev. D **57,** 630 (1998).

[29] A. S. Sengupta, S. V. Dhurandhar, A. Lazzarini, and T. Prince, Classical Quantum Gravity **19,** 1507 (2002).

[30] A. S. Sengupta, S. Dhurandhar, and A. Lazzarini, Phys. Rev. D **67,** 082004 (2003).

[31] K. Soni, B. U. Gadre, S. Mitra, and S. Dhurandhar, Phys. Rev. D **105,** 064005 (2022).

[32] B. P. Abbott *et al.*, Phys. Rev. X **9,** 031040 (2019).

[33] S. Roy, A. S. Sengupta, and N. Thakor, Phys. Rev. D **95,** 104045 (2017).

[34] S. Roy, A. S. Sengupta, and P. Ajith, Phys. Rev. D **99,** 024048 (2019).

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

[36] A. H. Nitz, Classical Quantum Gravity **35,** 035016 (2018).

[37] A. H. Nitz, T. Dent, T. D. Canton, S. Fairhurst, and D. A. Brown, Astrophys. J. **849,** 118 (2017).

[38] G. S. C. Davies and I. W. Harry, Classical Quantum Gravity **39,** 215012 (2022).

[39] C. Capano, T. Dent, C. Hanna, M. Hendry, C. Messenger, Y.-M. Hu, and J. Veitch, Phys. Rev. D **96,** 082002 (2017).

[40] D. M. Macleod, D. Meacher *et al.*, GWDataFind, https://git .ligo.org/computing/gwdatafind.

[41] D. Davis, J. S. Areeda, B. K. Berger, B. R., A. Effler, R. C. Essick, R. P. Fisher, P. Godwin, E. Goetz *et al.*, Classical Quantum Gravity **38,** 135014 (2021).

[42] V. Tiwari, Classical Quantum Gravity **35,** 145009 (2018).

[43] R. Sturani, S. Fischetti, L. Cadonati, G. M. Guidi, J. Healy, D. Shoemaker, and A. Vicere, J. Phys. Conf. Ser. **243,** 012007 (2010).

[44] C. Devine, Z. B. Etienne, and S. T. McWilliams, Classical Quantum Gravity **33,** 125025 (2016).

[45] J. W. Cooley and J. W. Tukey, Math. Comput. **19,** 297 (1965).

[46] LIGO Scientific Collaboration, LIGO algorithm library— LALSuite, free software (GPL), 10.7935/GT1W-FZ16 (2018).

[47] S. v. d. Walt, S. C. Colbert, and G. Varoquaux, Comput. Sci. Eng. **13** (2011).

[48] P. Virtanen *et al.*, Nat. Methods **17** (2020).

[49] A. M. Price-Whelan *et al.*, Astron. J. **156,** 123 (2018).