Template bank for compact binary coalescence searches in gravitational wave data: A general geometric placement algorithm

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

[1]*Department of Physics, Princeton University, Princeton, New Jersey 08540, USA*
[2]*School of Natural Sciences, Institute for Advanced Study,*
1 Einstein Drive, Princeton, New Jersey 08540, USA

We introduce an algorithm for placing template waveforms for the search of compact binary mergers in gravitational wave interferometer data. We exploit the smooth dependence of the amplitude and unwrapped phase of the frequency-domain waveform on the parameters of the binary. We group waveforms with similar amplitude profiles and perform a singular value decomposition of the phase profiles to obtain an orthonormal basis for the phase functions. The leading basis functions span a lower-dimensional linear space in which the unwrapped phase of any physical waveform is well approximated. The optimal template placement is given by a regular grid in the space of linear coefficients. The algorithm is applicable to any frequency-domain waveform model and detector sensitivity curve. It is computationally efficient and requires little tuning. Applying this method, we construct a set of template banks suitable for the search of aligned-spin binary neutron star, neutron-star–black-hole (NSBH) and binary black hole mergers in LIGO-Virgo data.

I. INTRODUCTION

The optimal algorithm to search for known signals in the presence of Gaussian noise is matched filtering, in which a signal template is cross-correlated with the data and triggers are recorded whenever the correlation exceeds some threshold. In the context of gravitational wave detection with the LIGO [1] and Virgo [2] interferometers, compact binary coalescences are a good example of predictable signals for which we have accurate models, and thus are well suited for matched filtering [3–5]. Indeed, the LIGO and Virgo collaborations have reported gravitational wave signals from ten binary black hole (BBH) and one binary neutron star (BNS) mergers during their first and second observing runs [6–13], all of which were found by search pipelines based on matched filtering [14,15] (seven of these BBHs were also found by an unmodeled search [6,16]). Searches in the public LIGO-Virgo data by independent groups have found seven additional BBHs [17,18], additional BBH candidates [19] and a BNS candidate [20], also employing matched filtering.

Since the source parameters describing the waveform are not known *a priori*, one needs a bank of waveform templates that adequately cover the parameter space. The notion of good coverage is characterized by the warranty that any physical waveform the search aims to detect has a sufficiently large match with at least one waveform in the

bank. For example, the LIGO and Virgo collaborations have aimed at a minimum match of 97% for any aligned-spin binary merger with component masses between 1 and $\sim 200\, M_\odot$, for which they require a template bank consisting of $\sim 4 \times 10^5$ waveforms [21]. Due to the large number of templates involved, matched filtering is a sizable computational task. This means that an efficient bank should not overcover the parameter space. In other words, the templates should be uniformly spaced with respect to a distance defined in terms of the matched-filtering mismatch between templates (defined in Sec. II). This incorporates the notion that, from the perspective of signal detection, two waveforms that sufficiently resemble each other are essentially indistinguishable in the presence of noise. Source parameters can be mutually degenerate in the sense that different parameter combinations may describe similar waveforms. The optimal placement of templates in physical parameter space is very nonuniform; for example, an order of magnitude more templates are needed to search for mergers with $1 - 3\, M_\odot$ components ("neutron stars") than for mergers with $3 - 200\, M_\odot$ ("black holes").

Two broad classes of template-placement algorithms have been developed in the literature. One robust method is "stochastic placement" [22–25]: waveforms are randomly drawn from the desired parameter space, and one gradually builds up the bank by only accepting newly drawn waveforms that differ sufficiently from the ones the bank already has, and rejecting those that are too similar to at least one existing waveform. Stochastic placement,

*jroulet@princeton.edu

however, has the shortcoming that a large number of trial waveforms needs to be drawn before convergence is achieved (much more than the required number of templates in the bank). This method also tends to overcover the parameter space, in the sense that the average template density is higher than optimal at fixed minimum match [26].

A different method to construct the bank is "geometric placement." Here, a metric in the parameter space is defined based on the matched-filtering overlap between waveforms [27,28]. This metric is then used to define a regular lattice [29–31]. However, it is in general difficult to derive this metric, especially if the parameter space is high dimensional or if the waveform model is not analytic. Approximations to the metric have first been found by using suitably reparametrized analytic, post-Newtonian (PN) nonspinning waveform models [27,28,32]; later generalizations include the use of phenomenological waveform models and template parameters [33], the inclusion of aligned-spin PN models [34,35], or numerical evaluation from arbitrary waveform models [36].

In practice, a combination of the two methods is often a better strategy. For example, one can place templates geometrically at low masses and stochastically at high masses [21,25], or one can use many small patches with regularly spaced templates, which are themselves placed stochastically to cover the entire parameter space [36].

In this work, we develop a fast and general method to construct a high-effectualness template bank using geometric placement. Our method relies on the construction of a flat, linear space of orthonormal phase functions that embeds the space of physical waveforms. The Euclidean distance in this space coincides with the mismatch distance between similar waveforms, making these coordinates naturally suited for geometric placement of templates. Besides optimal template placement, having this geometrical notion turns out to be helpful for a number of reasons. It allows to refine the bank locally around triggers at the time of search, reducing the amount of templates in the bank at fixed effective coverage. Moreover, a crucial stage of searches involves signal consistency checks that assess the probability that the residual between a best-fitting template and a candidate signal is explained by Gaussian noise in order to reject non-Gaussian noise transients [14,15,37–39]. With the bank described here, these tests can be made orthogonal to the linear space of waveforms, so that they are insensitive to mismatches due to the discreteness of the bank. This allows to make the tests more stringent and improves the sensitivity of the search [39]. We further require that the template bank be built from subbanks that can be approximated to have a fixed amplitude profile $A(f)$. This feature is useful for implementing the noise amplitude-spectral-density drift correction, a key component for precise matched filtering [38,40]. Together, these analytical properties make our template

bank appealing, even considering that there are other template banks with comparable effectualness and number of templates in the literature. Finally, building a new template bank enables us to customize a number of other choices, like the frequency range and parameter space covered, in the context of our search pipeline [38] and the detector performances during the observing time analyzed. The coordinates presented in this work are similar in essence to the ones introduced by Brown *et al.* [34], except that we generalize them to arbitrary waveform models and component mass ranges.

The paper is organized as follows. In Sec. II we define a metric based on the mismatch between templates and show how the desired Euclidean space can be constructed. In Sec. III we apply this formalism to the construction of a search-quality template bank that targets stellar-mass compact binary mergers. We summarize our results in Sec. IV. The bank presented here was used in the searches described in Refs. [18,38], except that the bank used in [38] had some limited differences that we report in the Appendix.

II. LINEAR METRIC SPACE

In this section, we define the notion of distance between templates and describe the construction of a low-dimensional linear space of phase functions in which the metric is Euclidean. We build this linear space based on the intuition that the unwrapped phases are smooth functions of the wave frequency [41] and hence are linear combinations of a small number of basis functions [32,34].

A. Mismatch distance

We first introduce the noise-weighted inner product in the frequency domain [5]

$$(h_i|h_j) := 4 \int_0^\infty \mathrm{d}f \, \frac{\tilde{h}_i(f)\tilde{h}_j^*(f)}{S_n(f)}. \tag{1}$$

Here, $S_n(f)$ is a fiducial one-sided power spectral density (PSD) of the detector noise and tildes indicate Fourier transforms. The match between h_i and h_j is given by $\Re z_{ij}$, where

$$z_{ij} := \frac{(h_i|h_j)}{\sqrt{(h_i|h_i)(h_j|h_j)}} \equiv (h_i|h_j). \tag{2}$$

In the second line, we normalize the waveforms to

$$(h|h) = 1, \tag{3}$$

as usually the template waveforms are defined up to an overall normalization. Since all possible coalescence times and phases are searched for, waveforms related by time and phase offsets are described by the same waveform in the template bank. Thus, the match is maximized over time and phase offsets:

$$m_{ij} := \max_{\tau_0, \phi_0}(\Re z_{ij}(\tau_0, \phi_0)) = \max_{\tau_0}|z_{ij}(\tau_0)|, \qquad (4)$$

where τ_0 and ϕ_0 are the time and phase offsets between h_i and h_j, respectively. We define the mismatch distance d_{ij} between the two waveforms by

$$d_{ij}^2 = 1 - m_{ij}. \qquad (5)$$

We seek a parametrization of waveforms under which the mismatch distance has a Euclidean metric for similar waveforms.

B. Linear space

A general frequency-domain waveform model can be cast to the form

$$\tilde{h}(f;\boldsymbol{p}) = A(f;\boldsymbol{p})e^{i\psi(f;\boldsymbol{p})}. \qquad (6)$$

Under the approximation that the dominant mode of gravitational radiation has $(\ell, |m|) = (2, 2)$ and that spin-orbital precession and eccentricity effects are insignificant, the frequency dependent functions A and ψ vary slowly with the binary parameters $\boldsymbol{p}$, as illustrated in Fig. 1. For matched filtering, the phase ψ is the most important to describe with high accuracy, since loss of phase coherence leads to a rapid degradation of the signal-to-noise ratio (SNR).

Moreover, it is important to analyze templates with different amplitude profiles $A(f)$ separately as the matched filtering correction from PSD drifts depends on $A(f)$ [40]. Thus we assume in the following that $A(f;\boldsymbol{p}) \approx \bar{A}(f)$ is valid for a suitably chosen subset of parameters. To achieve
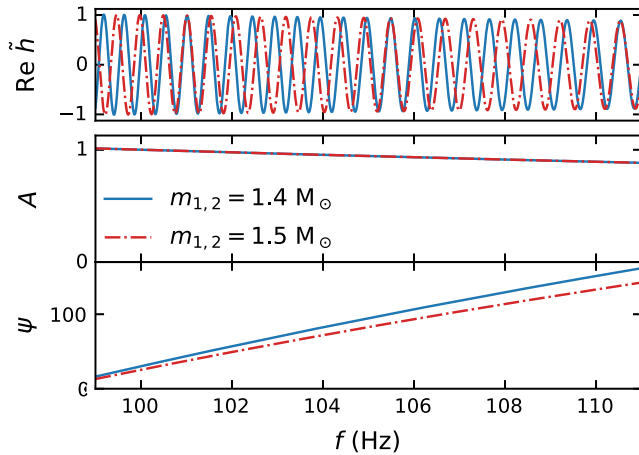

FIG. 1. An example of two waveforms that look very different to each other in the frequency domain (top panel) but have very similar amplitude and phase profiles (middle and bottom panels). The amplitude and phase profiles can be well captured by a low-dimensional linear space spanned by a few basis functions. Waveform amplitudes are shown in arbitrary units.

this, we sort a large number of randomly sampled physical input waveforms into groups of similar amplitude profile. In each group, we require that the match of the amplitudes to a reference $\bar{A}(f)$ exceeds a minimum

$$(A_i|\bar{A}) = 4 \int_0^\infty df \frac{A_i(f)\bar{A}(f)}{S_n(f)} \geq 0.96 \qquad (7)$$

for all input waveforms h_i in the group. Note that the match of the amplitudes sets an upper bound on the match of the waveforms. Our approach will be to split a template bank into "subbanks," each subbank describing one group of input waveforms which share the same approximate amplitude profile $\bar{A}(f)$.

We design the subbanks in order to minimize the average amplitude mismatch as follows. We start with a single subbank that contains all the input waveforms, and define its reference amplitude profile as the root-mean-square

$$\bar{A}(f) := \sqrt{\langle A^2 \rangle(f)}, \qquad (8)$$

where the angled brackets indicate average over the input waveforms in the subbank. This choice inherits the normalization of the input waveforms. We compute the amplitude match Eq. (7) for all the waveforms; if the worst match satisfies the chosen bound we stop. If it does not, we add a new subbank with a reference amplitude given by the waveform with the worst amplitude match. We then optimize the choice of reference amplitudes using the k-means algorithm: we reassign waveforms to subbanks by their best amplitude match, redefine the amplitude profile of the subbanks using Eq. (8), and iterate these two steps a few times to achieve convergence. Finally we recompute the worst match and decide if a new subbank is needed, in which case we repeat the above process.

Having decided on the division of subbanks, we wish to find an efficient representation of the set of phases $\psi(f)$ as a linear combination of a small number of basis functions,

$$\psi(f;\boldsymbol{p}) = \bar{\psi}(f) + \sum_\alpha^{\text{few}} c_\alpha(\boldsymbol{p})\psi_\alpha(f), \qquad (9)$$

where α is an integer index that enumerates the basis functions and $\bar{\psi}(f)$ is an average phase which we are free to define. From now on, we abandon the physical parameters $\boldsymbol{p}$ and describe the waveforms in terms of their c_α components:

$$h(f;\boldsymbol{c}) = \bar{A}(f)\exp\left[i\left(\bar{\psi}(f) + \sum_\alpha c_\alpha \psi_\alpha(f)\right)\right]. \qquad (10)$$

We now express the match between two waveforms using the above decomposition. As mentioned earlier, template waveforms are defined up to arbitrary time and

phase offsets, namely an additive piece to the phase that is a linear function of the frequency $\Delta\psi(f) = \phi_0 + 2\pi f\tau_0$. We choose the first two basis functions $\psi_0(f)$ and $\psi_1(f)$ to span the subspace of linear phases so that c_0 and c_1 capture phase and time offsets, respectively, and in particular $\psi_0(f) \equiv 1$. If two waveforms are similar, their inner product Eq. (1) to second order in δc_α is approximately

$$(h(\mathbf{c})|h(\mathbf{c}+\boldsymbol{\delta c})) = 4\int_0^\infty df \frac{\bar{A}^2(f)}{S_n(f)} e^{i\sum_\alpha \delta c_\alpha \psi_\alpha(f)}$$

$$\approx 4e^{i\delta c_0}\int_0^\infty df \frac{\bar{A}^2(f)}{S_n(f)}\left[1 + i\sum_{\alpha\geq 1}\delta c_\alpha\psi_\alpha(f)\right.$$

$$\left. - \frac{1}{2}\sum_{\alpha,\beta\geq 1}\delta c_\alpha\delta c_\beta\psi_\alpha(f)\psi_\beta(f)\right] + \mathcal{O}(\delta c^3).$$

$$(11)$$

This motivates a new inner product, with respect to which we will orthonormalize the basis functions:

$$\langle\psi_\alpha,\psi_\beta\rangle := 4\int_0^\infty df \frac{\bar{A}^2(f)}{S_n(f)}\psi_\alpha(f)\psi_\beta(f) \equiv \delta_{\alpha\beta}, \quad (12)$$

which we enforce by a suitable choice of the basis functions $\psi_\alpha(f)$ (described below). In particular, the first condition $\langle\psi_0,\psi_0\rangle = 1$ is the normalization Eq. (3), and the two first basis functions are

$$\psi_0(f) = 1,$$

$$\psi_1(f) = \frac{f - \bar{f}}{\sqrt{\bar{f^2} - \bar{f}^2}}, \quad (13)$$

where we define $\bar{f}^n := 4\int_0^\infty df f^n A^2(f)/S_n(f)$.

Using orthonormality, Eq. (11) becomes

$$(h(\mathbf{c})|h(\mathbf{c}+\boldsymbol{\delta c})) \approx e^{i\delta c_0}\left(1 - \frac{1}{2}\sum_{\alpha\geq 1}\delta c_\alpha^2\right) + \mathcal{O}(\delta c^3). \quad (14)$$

Thus, for nearby templates the distance Eq. (5) is

$$d_{\mathbf{c},\mathbf{c}+\boldsymbol{\delta c}}^2 \approx \frac{1}{2}\sum_{\alpha\geq 2}\delta c_\alpha^2 + \mathcal{O}(\delta c^3), \quad (15)$$

which means that the mismatch distance is given by a Euclidean metric in $\mathbf{c}$ space at small displacements. We construct the bank on a regular grid in $\mathbf{c}$ space with spacings $\Delta c_\alpha \lesssim 1$, chosen sufficiently small so as to guarantee a minimal loss of match.

We note in passing that we can also compute the distance in the opposite limit of large separation, which is useful for estimating the long-range correlations between triggers from different templates during a search. Assuming now

that the templates are separated by $\boldsymbol{\delta c} = D\hat{\mathbf{n}}$, with $\sum_\alpha n_\alpha^2 = 1$ and $D \gg 1$, we can perform a stationary phase approximation around the frequencies f_j at which $\sum_\alpha n_\alpha\psi'_\alpha(f_j) = 0$. This yields

$$(h(\mathbf{c})|h(\mathbf{c}+D\hat{\mathbf{n}}))$$

$$= 4\int_0^\infty df \frac{\bar{A}^2(f)}{S_n(f)} e^{iD\sum_\alpha n_\alpha\psi_\alpha(f)}$$

$$\approx \frac{4}{\sqrt{D}}\sum_j \frac{\bar{A}^2(f_j)}{S_n(f_j)}\frac{\sqrt{2\pi}\exp(-i\frac{\pi}{4}+iD\sum_\alpha n_\alpha\psi_\alpha(f_j))}{\sqrt{\sum_\alpha n_\alpha\psi''_\alpha(f_j)}}.$$

$$(16)$$

Thus, the long-range correlation between two templates separated by D decays as $1/\sqrt{D}$ (this holds for the match without maximization over time).

In practice we choose the set of basis functions $\psi_\alpha(f)$ as follows:

(1) Define a discrete frequency grid $\{f_k\}$ (our choice is described in Sec. III). The integrals over frequency will be approximated by quadratures $\sum_k \Delta f_k\dots$.

(2) Compute a moderately large number of waveforms for random parameter choices (we use 5×10^4), and extract the unwrapped phases, $\{\psi^{(i)}(f_k)\}$, as illustrated in the top panel of Fig. 2.
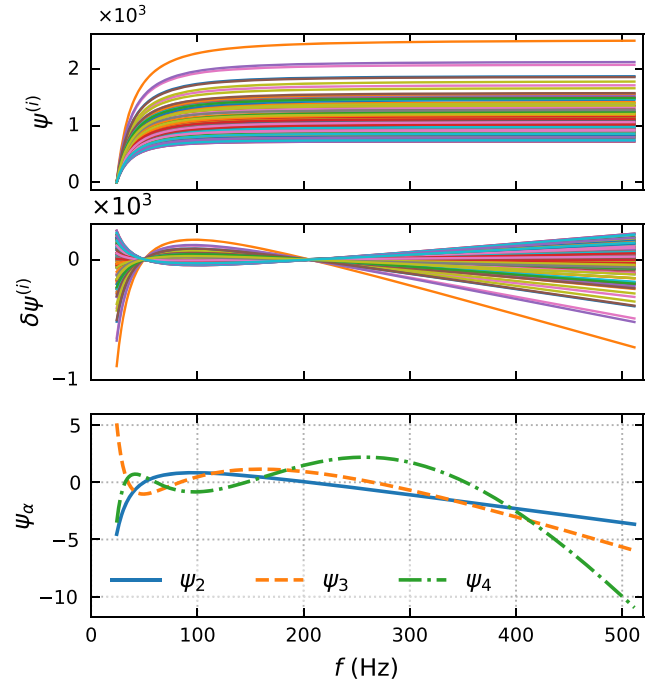
(3) Subtract the average phase $\bar{\psi}(f)$.

FIG. 2. Construction of the basis functions ψ_α. Top panel: (A subset of 100) input unwrapped phases for random parameters. Middle panel: Phase residuals after subtraction of the average phase and orthogonalization with respect to time and phase offsets. Bottom panel: First three basis functions.

(4) Subtract the projection onto the first two dimensions so that the phase residuals

$$\delta\psi^{(i)}(f) = \psi^{(i)}(f) - \bar{\psi}(f)$$
$$- \sum_{\alpha=0}^{1} \langle \psi^{(i)} - \bar{\psi}, \psi_\alpha \rangle \psi_\alpha(f) \quad (17)$$

are orthogonal to ψ_0 and $\psi_1(f)$ with respect to the inner product Eq. (12) (middle panel of Fig. 2).

(5) Construct a matrix of weighted phase residuals

$$M_{ik} = w_k \delta\psi^{(i)}(f_k),$$
$$w_k = 2\bar{A}(f_k)\sqrt{\Delta f_k/S_n(f_k)}, \quad (18)$$

and find its singular-value decomposition (SVD)

$$M_{ik} = \sum_\alpha U_{i\alpha} D_\alpha V_{\alpha k}. \quad (19)$$

U, V are orthogonal matrices and we sort the axes so that the eigenvalues $D_\alpha > 0$ are in decreasing order. From the orthogonality of V, i.e., $\sum_k V_{\alpha k} V_{\beta k} = \delta_{\alpha\beta}$, we can identify

$$V_{\alpha k} = w_k \psi_\alpha(f_k) \quad (20)$$

which satisfies the orthonormality Eq. (12) and defines the basis functions, with the convention that the α start at 2 (bottom panel of Fig. 2).

From Eqs. (9) and (19) it follows that the components of the input waveforms are

$$c_\alpha^{(i)} = U_{i\alpha} D_\alpha. \quad (21)$$

Since U is an orthogonal matrix, $|U_{i\alpha}| \leq 1$ and $|c_\alpha| \leq D_\alpha$, that is, the extent spanned by the input samples along each dimension in component space is bounded by D_α. This means that the information in the templates is captured by the first few components along the larger dimensions, and we can reduce the dimensionality of our description by dropping the dimensions that have $D_\alpha \ll 1$.

III. CONSTRUCTING A SEARCH QUALITY TEMPLATE BANK

In this section, we apply the method developed in Sec. II to the construction of a template bank suitable to the search of gravitational wave strain signals from binary neutron stars, neutron-star–black-hole and binary black hole mergers.

We choose lower and upper frequency cutoffs of $f_{\min} = 24$ Hz, and $f_{\max} = 512$ Hz, respectively. These cutoffs are chosen such that the resulting loss in SNR^2 is lower than 2% for binary neutron star templates (the amplitude profiles of these whitened waveforms, i.e., $A(f)/\sqrt{S_n(f)}$, are essentially independent of parameters since the cutoff scale is outside the LIGO sensitivity band). Formally, the accumulated SNR$^2 = 4 \int_0^{24 \text{ Hz}} df A^2(f)/S_n(f) \approx 4 \int_{512 \text{ Hz}}^\infty df A^2(f)/S_n(f) \approx 10^{-2}$ outside our frequency range. It is advisable to restrict the frequency range because the linear-free phase, and thus the basis functions, grow rapidly at both ends (see Fig. 2), and our Taylor expansion Eq. (11) would become inaccurate. As we noted above, it is exactly at these frequencies where the contribution to the matched-filtering SNR vanishes. It is better to discard these frequencies rather than to try and capture the negligible information content within by adding extra dimensions to the template bank. Furthermore, this has the additional benefit that the strain data can be downsampled during analysis, which reduces the computational cost of the search.

We define the fiducial PSD empirically from the PSDs of 200 LIGO Handford and LIGO Livingston data files chosen randomly from the Second Advanced LIGO Observing Run (O2) release [42,43]. Each individual PSD was computed as described in [38]. The fiducial PSD is constructed using the tenth percentile of all the sample PSDs in each frequency bin. This choice is robust to large fluctuations in the sample PSDs, and is representative of optimal detector conditions.

We choose a target parameter space of compact binary mergers satisfying the following bounds:

$$1M_\odot < m_2 < m_1 < 100M_\odot \quad (22)$$

$$q > \begin{cases} 1/50 & \text{if } m_2 < 3M_\odot \\ 1/18 & \text{otherwise,} \end{cases} \quad (23)$$

$$|\chi_1|, |\chi_2| < 0.99, \quad (24)$$

where m_1 and m_2 are the primary and secondary masses, respectively, $q = m_2/m_1 \leq 1$ is the mass ratio, and χ_1 and χ_2 are the individual dimensionless spin projections in the direction of the orbital angular momentum. The parameter ranges and approximant used are not a constraint from the LIGO and Virgo detectors or the method presented here, but a documentation of the choices we made. In particular, the mass ratio cut for BBHs was due to the calibration regime of the IMRPHENOMD approximant [44]. For NSBH, we extend the maximal q because a substantial part of the NSBH parameter space lies outside the calibrated range of IMRPHENOMD. For the purpose of signal detection (as opposed to parameter estimation), the calibration tolerance is less stringent, as long as a signal can be recovered by the model with some combination of parameters. Compared to other template banks in the literature, the one presented here covers a larger spin range for low-mass objects. Indeed, bounds of $|\chi| < 0.05$ [21,34,36] or $|\chi| < 0.4$ [34] have been used in the BNS mass range, the

former motivated by the known binary neutron star spins and the latter by the known pulsar spins [45]. Neutron stars can in principle have dimensionless spins up to a mass-shedding limit of $|\chi| \sim 0.7$ [46,47]. Other types of compact objects, including light black holes, may in principle have even higher spins. This motivates us to cover this unexplored part of the parameter space.

As mentioned before, the number of templates required to describe waveforms from low-mass mergers is significantly larger than that for high-mass mergers, due to the larger number of wave cycles in band. Searches with larger template banks suffer a penalty in sensitivity because of the increased look-elsewhere effect. To prevent the high penalty inherent to the lower-mass region of parameter space from affecting the higher-mass regions, we propose to divide the search space into a number of regions and perform an independent search in each. Each search then only pays an additional look-elsewhere penalty that a few other searches are performed, but is unaffected by the potentially huge size of the other banks. This division can be interpreted as implementing a prior about which templates are more likely to produce an astrophysical trigger: if we expect comparable numbers of high- and low-mass signals but have vastly more templates at low mass, any particular low-mass template is much less likely to produce an astrophysical trigger. In addition, templates in different regions of parameter space are sensitive to different types of noise transients in the strain data. Dividing the search into several regions enables us to recognize the different types of noise background that a search using each class of templates is subject to.

Under the above motivations, we divide the search space into regions based on the component masses, and construct a separate template bank for each of them. The division is illustrated in Fig. 3 and is defined as follows. We refer to binary components with masses between 1 and 3 $M_\odot$ as neutron stars, and to components with masses between 3 and 100 $M_\odot$ as black holes. We make three binary neutron star template banks, three neutron-star–black-hole (NSBH) banks, and five binary black hole banks. The banks within each of these categories are defined by bins in the chirp mass $\mathcal{M} := (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$. We put the bounds between the three BNS banks at $\mathcal{M} = \{1.1, 1.3\} M_\odot$. This choice is motivated by the observation that the chirp masses of the known Galactic binary neutron stars expected to merge within a Hubble time lie in a narrow range [48], and therefore we might expect more astrophysical signals from this chirp mass range (which we further expand to account for the redshift of the detector-frame masses up to $z \sim 0.05$, or a luminosity distance $d_L \sim 200$ Mpc). In this way, we minimize the number of templates in the most astrophysically probable BNS bank, BNS 1, enhancing our sensitivity to those systems. A similar strategy was adopted in Refs. [20,49]. For other banks, we use logarithmic chirp-mass bins: we place the bounds between the three NSBH
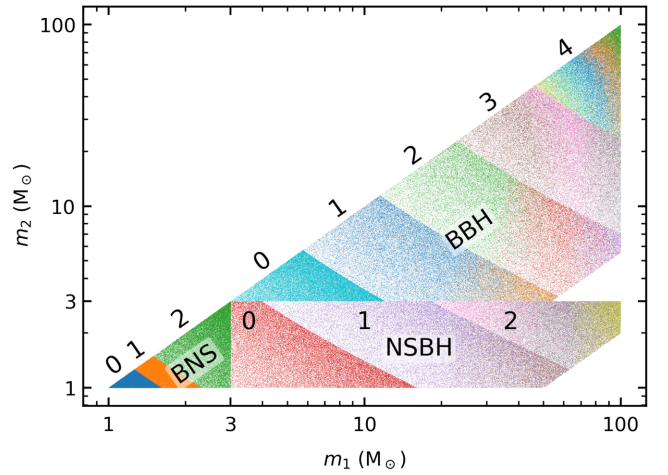

FIG. 3.   Component masses of the input waveforms used for constructing our template banks. We divide the parameter space according to the component masses into three banks of binary neutron star waveforms (BNS 0-2), three banks of neutron-star–black-hole waveforms (NSBH 0-2), and five banks of binary black hole waveforms (BBH 0-4). We further divide each bank into subbanks (color coded) under the criterion that the match of the amplitude profiles with a reference profile specific to each subbank exceeds 0.96. The dots show the 5×10^4 input waveforms we use to build each bank. We do not scatter plot here individual template waveforms in each bank, as those lack an explicit representation in terms of the source physical parameters.

banks at $\mathcal{M} = \{3, 6\} M_\odot$, and those between the five BBH banks at $\mathcal{M} = \{5, 10, 20, 40\} M_\odot$. We generate 5×10^4 input waveforms in each bank using the IMRPHENOMD approximant [44]. Based on the amplitude profiles $A_i(f)$ of the input waveforms, we further divide each bank into subbanks as explained in Sec. II. We find that a single subbank is sufficient for waveforms with $m_{1,2} \lesssim 15\ M_\odot$, but multiple amplitude subbanks are needed for heavier mergers as the frequency at which $A(f)$ is cut off falls within the LIGO sensitive band. Table I summarizes the parameters of all template banks. The banks differ greatly in size, which justifies the division of the search space into multiple banks.

For each subbank, we apply the procedure outlined in Sec. II to define a set of basis phase functions that generate a linear space and obtain the projections of the input waveforms onto that space. These are shown in Fig. 4 for the example case of the BBH 0 bank, with the points color coded by their chirp mass. The first three dimensions capture practically all the diversity of the input waveforms. Also note the large differences in size from the leading dimension to the subleading ones. The number of cycles, proportional to $\mathcal{M}^{-5/3}$, is the best-measured parameter and thus should approximately correspond to the coefficient of the leading dimension [4,41]. Indeed, this is observed in Fig. 4, confirming that the decomposition is working as expected.

TABLE I.   Summary of the parameters of the template banks. Columns 2 to 6 describe the bounds of physical parameter space that each bank is designed to cover. ζ is a tunable fudge parameter that controls the tolerance for removing nonphysical grid points. Δc_α is the grid spacing that we chose for each bank. $N_{subbanks}$ is the resulting number of subbanks in each bank. $d_{subbanks}$ and $L_{max,subbanks}$ are the dimensionalities of each subbank (sorted by increasing mean total mass) and the size of their largest dimension, respectively. $N_{templates}$ is the total number of templates in each bank.

Bank	$m_1(M_\odot)$	$m_2(M_\odot)$	$\mathcal{M}(M_\odot)$	q_{min}	$|\chi_{1,2}|_{max}$	ζ	Δc_α	$N_{subbanks}$	$d_{subbanks}$	$L_{max,subbanks}$	$N_{templates}$
BNS 0			<1.1					1	2	777.0	48806
BNS 1	(1,3)	(1,3)	(1.1,1.3)	⋯	0.99	0.05	0.55	1	2	434.3	23856
BNS 2			>1.3					1	2	824.6	43781
NSBH 0			<3					1	4	753.4	84641
NSBH 1	(3,100)	(1,3)	(3,6)	1/50	0.99	0.05	0.5	2	6,6	259.5,166.8	85149
NSBH 2			>6					3	5,4,4	87.5,61.2,9.4	15628
BBH 0			<5				0.55	1	3	270.6	8246
BBH 1			(5,10)				0.55	2	4,4	113.7,50.0	4277
BBH 2	(3,100)	(3,100)	(10,20)	1/18	0.99	0.05	0.5	3	3,4,3	41.5,33.5,10.3	1607
BBH 3			(20,40)				0.45	3	2,2,2	11.7,10.8,4.9	225
BBH 4			>40				0.35	5	2,2,2,1,1	2.9,2.0,1.1,0.7,0.5	46
Total											316262

Next, we choose a grid spacing Δc_α common to all dimensions and define a rectangular grid in component space as follows. We force the point $\boldsymbol{c} = \boldsymbol{0}$ to be a grid point, because the SVD typically aligns the highest density regions (where the input physical waveforms tend to be) with the axes. Along each dimension, we add uniformly spaced points until the whole range spanned by the input waveforms is covered. We allow the spacing to slightly decrease so that the most extreme input component is half the grid spacing away from the most extreme grid point. We do this for each dimension and in the positive and negative directions separately. Finally, not all the points of the rectangular grid describe physically viable waveforms. We only keep the templates that are close to at least one input
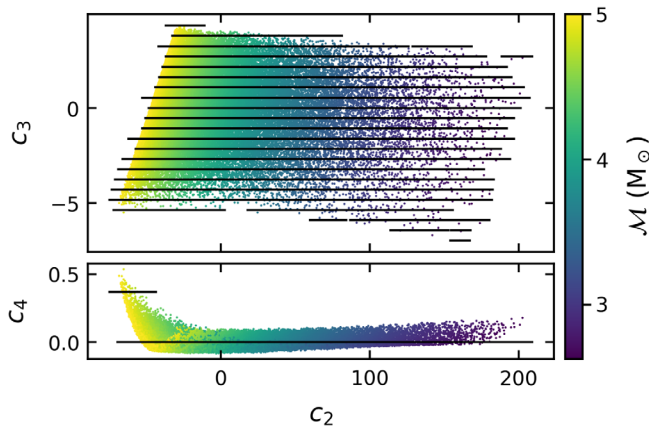

FIG. 4.   Our three-dimensional BBH 0 template bank in the component space (black dots, which at this scale appear as lines), projected along two different axes. Underlaid are the input waveforms used to build the bank, projected according to Eq. (21) and color coded by their chirp mass.

waveform, with the following criterion. For every input waveform set of components, we keep the closest grid point and a patch of the grid around it, with size equal to the corresponding dimension times a tunable fudge factor $\zeta \sim 0.1$.

Indeed, as Fig. 4 shows, the input physical waveforms do not fill the entire rectangular volume but are distributed within some irregularly shaped region. Furthermore, the density of input waveforms is low in the low-$\mathcal{M}$ region, where the waveforms have more wave cycles in band and hence are mutually more distinguishable. Holes can be produced in the physically viable region if the fudge factor ζ is too small, and there is an excess of unphysical templates if ζ is too large. We choose the Δc_α and ζ parameters such that we achieve a good balance between economic template bank size and high bank effectualness. The values chosen for each bank are reported in Table I.

In Table I we observe a general trend with the mass: the banks for lighter mergers tend to have fewer subbanks and the first dimension spans a wider range. By comparison, the banks for heavier mergers have more subbanks, with smaller dimensions. The increase in the number of subbanks for heavier mergers is caused by the cutoff frequency falling in the band, which increases the variety of amplitude profiles.

There are interesting implications of the number of dimensions and their size for parameter estimation. Given an astrophysical signal, in the limit of high SNR ρ, the parameter likelihood is approximately given by $P(d|\boldsymbol{p}) \propto \exp(-\rho^2 |z|^2/2)$, where $z = (h(\boldsymbol{p}_*)|h(\boldsymbol{p}))$ is the complex match of $\boldsymbol{p}$ to the best-fit parameters $\boldsymbol{p}_*$ [50]. By virtue of Eq. (15), this means that the likelihood is approximately an isotropic Gaussian in terms of the c_α coordinates, with a width $\sim 1/\rho$. The number of dimensions

can therefore be interpreted as the number of independent parameters that can be measured, and the size of each dimension as the relative precision that can be obtained for a fixed SNR (with the caveat that we have restricted the frequency range; for example, information about the tidal deformability comes from frequencies higher than our 512 Hz cutoff).

For example, for BNS (and effectively for light BBH) the banks have two dimensions, with a large first dimension well correlated with the chirp mass (Fig. 4). The two measurable parameters are the chirp mass, which indeed can be measured to much higher precision than for heavy systems, and a combination of the mass ratio and effective spin which can be measured with a lower precision. These are the leading contributions to the phase evolution as can be understood from the post-Newtonian expansion.

An important advantage of our geometric coordinates is that they are well suited for a two-step search that effectively achieves a smaller grid spacing at reduced computational cost. We realize this by refining the template grid on demand around all triggers that exceed an appropriately lowered SNR threshold [38,51]. During the search, we first use a coarse grid, and refine every trigger using neighboring templates from a denser grid that has half the spacing along each dimension. The fact that the distance between c_α components translates directly to mismatch [Eq. (15)] makes this method straightforward to implement.

To characterize the effectualness of the bank at recovering the target physical signals, we generate a set of 10^4 random "test waveforms" within the parameter range of each bank, using the same approximant with which the input waveforms were generated. We choose the parameters from a distribution that is uniform in the component masses m_1, m_2 and aligned spins χ_1, χ_2. In principle, we would have to match each test waveform against every waveform in the bank to look for the best match. To save computational effort, we select a candidate best match based on the approximate metric Eq. (15) by extracting the phase of the test waveform $\psi^{(i)}(f)$, projecting it onto the linear space, $c_\alpha^{(i)} = \langle \psi^{(i)} - \bar\psi, \psi_\alpha \rangle$, and finding the closest grid point with respect to the Euclidean metric (15). Since *a priori* we do not know which subbank best describes the test waveform, we pick the best candidate from each subbank and compute the match with all. The best match with our reduced set of candidates is a lower bound on the best match over all the waveforms in the bank. Rather than using Eq. (4) directly, we compute the match by following the detection strategy described in Venumadhav *et al.* [38]: we account for the finite time resolution of the Fourier transform by downsampling the waveforms to 512 Hz and sinc-interpolating the matched-filter output twice. We show the result of this test in Fig. 5, in terms of the cumulative fraction of the matches with each bank before and after applying the grid refinement, which we use to assess the collection threshold on the coarse grid and the effectualness
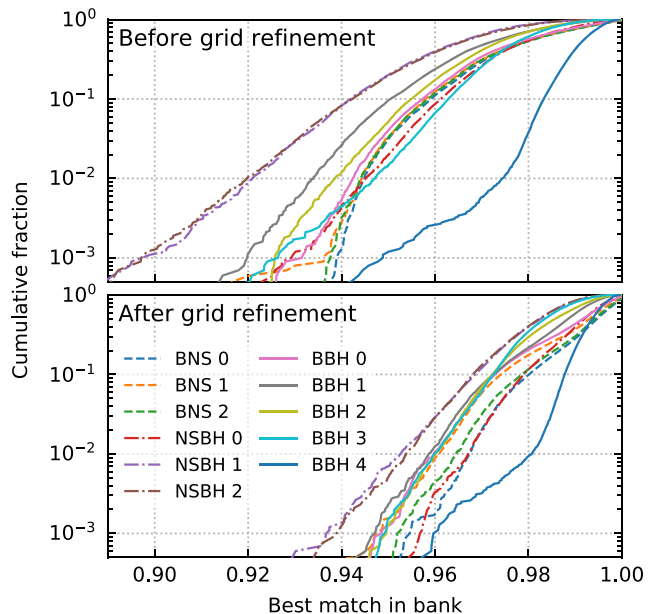

FIG. 5. Effectualness of our template banks, tested on random waveforms drawn from a distribution uniform in individual masses and aligned spins. The vertical axis shows the fraction of the random trials that do not achieve a given match in the bank.

achieved for each bank, respectively. We find that depending on the bank 99% of the templates have a match higher than 0.95 to 0.98.

IV. CONCLUSIONS

We have developed a general and computationally efficient geometric placement algorithm to construct high-effectualness template banks for detecting gravitational waves from compact binary mergers. We have constructed a basis of functions that generate a linear space of phase profiles on which the mismatch metric is Euclidean. For the purpose of signal detection, we shift the focus away from physical parameters to the linear coefficients for the basis phase profiles. We identified which components carry the largest amount of information about physical waveforms and what is the minimal set required to guarantee a desired match. The basis functions can be determined from a set of input waveforms whose size is small compared to that of the bank. The basis functions can be generated with any frequency-domain waveform model. The resolution of the bank can be decided independently after the basis functions have been found; in particular, it can be increased arbitrarily at negligible computational cost since no further evaluations of the physical waveform approximants need to be done. Our algorithm guarantees that within each of the few subbanks that make up one template bank, all templates share the same amplitude profile, a property that is critical for the correction of the power-spectral-density drift in signal processing.

We have applied our algorithm to the construction of a collection of eleven template banks that together cover the

parameter space associated to stellar-mass compact binary mergers with aligned spins. We find the effectualness and total number of templates to be comparable to the ones obtained by other algorithms in the literature [21,36]; detailed comparisons are difficult due to the different parameter spaces targeted in various works. We note that our template bank includes rapidly spinning neutron stars, which to date have not been searched for in the gravitational wave data. We implement a two-step search with a coarse grid that we refine around triggers at the time of search, a task for which our new formalism is ideally suited. This is an important step to reduce the number of templates while preserving a high effectualness.

Looking forward, an accurate and fast interpolation from physical parameters to the c_α component space would be extremely useful for rapid parameter estimation. Firstly, because waveforms can be generated at negligible computational cost once the components are known. At least in cases where analytical waveform models are not valid, waveform generation dominates the computational cost of parameter estimation. Moreover, the likelihood would look close to an isotropic Gaussian in terms of the c_α coordinates due to orthonormality, making them a suitable choice from the data analysis perspective. Other natural extensions of the work presented here are to include the effects of precession, due to misalignment between the spins and the orbital angular momentum, and eccentricity. These are deferred for future work. The inclusion of eccentricity is currently limited by the availability of robust public waveform generation codes.

The template bank described here is publicly available.[1]

ACKNOWLEDGMENTS

APPENDIX: DIFFERENCES WITH THE BANK USED BY VENUMADHAV *et al.* [38]

In this Appendix we report the differences between the template bank presented in this work and the one that was actually used in the binary black hole search reported by Venumadhav *et al.* [38].

(i) The bank used in Ref. [38] was restricted to BBHs, with a narrower range for the aligned spins $|\chi_{1,2}| < 0.85$ instead of 0.99.

(ii) The reference PSD used to build the bank was `aLIGO_MID_LOW` [52] instead of an empirically measured one.

(iii) As a consequence, the bank used a low-frequency cutoff at 20 Hz instead of 24 Hz. We found that with the empirical PSD, the relative contribution to squared SNR from frequencies below 24 Hz is $<1\%$.

(iv) The optimization of the subbank amplitude profiles with the k-means clustering algorithm was not done; instead the division into subbanks was computed in a way analogous to the stochastic placement approach to building a template bank described in Sec. I, with a required amplitude match $(A_i|\bar{A}) > 0.95$. The reference amplitude of the subbank was given by the first waveform generated in the subbank. Using the k-means clustering improves the best match by ~ 0.01 on average.

---

[1]https://github.com/jroulet/template_bank.
[2]https://www.gw-openscience.org.

---

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

[2] F. Acernese *et al.*, Classical Quantum Gravity **32,** 024001 (2015).

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

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

[5] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, Phys. Rev. D **85,** 122006 (2012).

[6] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), arXiv:1811.12907.

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

[8] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **116**, 241103 (2016).

[9] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. X **6**, 041015 (2016).

[10] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **118**, 221101 (2017).

[11] B. P. Abbott *et al.*, Astrophys. J. **851**, L35 (2017).

[12] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **119**, 141101 (2017).

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

[14] S. Sachdev, S. Caudill, H. Fong, R. K. L. Lo, C. Messick, D. Mukherjee, R. Magee, L. Tsukada, K. Blackburn *et al.*, arXiv:1901.08580v1.

[15] 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 *et al.*, Classical Quantum Gravity **33**, 215004 (2016).

[16] S. Klimenko, G. Vedovato, M. Drago, F. Salemi, V. Tiwari, G. A. Prodi, C. Lazzaro, K. Ackley, S. Tiwari *et al.*, Phys. Rev. D **93**, 042004 (2016).

[17] B. Zackay, T. Venumadhav, L. Dai, J. Roulet, and M. Zaldarriaga, arXiv:1902.10331.

[18] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, arXiv:1904.07214.

[19] A. H. Nitz, C. Capano, A. B. Nielsen, S. Reyes, R. White, D. A. Brown, and B. Krishnan, Astrophys. J. **872**, 195 (2019).

[20] A. H. Nitz, A. B. Nielsen, and C. D. Capano, Astrophys. J. **876**, L4 (2019).

[21] T. Dal Canton and I. W. Harry, arXiv:1705.01845.

[22] I. W. Harry, B. Allen, and B. S. Sathyaprakash, Phys. Rev. D **80**, 104014 (2009).

[23] P. Ajith, N. Fotopoulos, S. Privitera, A. Neunzert, N. Mazumder, and A. J. Weinstein, Phys. Rev. D **89**, 084041 (2014).

[24] S. Privitera, S. R. P. Mohapatra, P. Ajith, K. Cannon, N. Fotopoulos, M. A. Frei, C. Hanna, A. J. Weinstein, and J. T. Whelan, Phys. Rev. D **89**, 024003 (2014).

[25] C. Capano, I. Harry, S. Privitera, and A. Buonanno, Phys. Rev. D **93**, 124007 (2016).

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

[27] B. J. Owen, Phys. Rev. D **53**, 6749 (1996).

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

[29] S. Babak, R. Balasubramanian, D. Churches, T. Cokelaer, and B. S. Sathyaprakash, Classical Quantum Gravity **23**, 5477 (2006).

[30] T. Cokelaer, Phys. Rev. D **76**, 102004 (2007).

[31] S. Babak, R. Biswas, P. R. Brady, D. A. Brown, K. Cannon, C. D. Capano, J. H. Clayton, T. Cokelaer, J. D. E. Creighton *et al.*, Phys. Rev. D **87**, 024033 (2013).

[32] T. Tanaka and H. Tagoshi, Phys. Rev. D **62**, 082001 (2000).

[33] P. Ajith, S. Babak, Y. Chen, M. Hewitson, B. Krishnan, A. M. Sintes, J. T. Whelan, B. Brügmann, P. Diener *et al.*, Phys. Rev. D **77**, 104017 (2008).

[34] D. A. Brown, I. Harry, A. Lundgren, and A. H. Nitz, Phys. Rev. D **86**, 084017 (2012).

[35] I. W. Harry, A. H. Nitz, D. A. Brown, A. P. Lundgren, E. Ochsner, and D. Keppel, Phys. Rev. D **89**, 024010 (2014).

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

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

[38] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, arXiv:1902.10341.

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

[40] B. Zackay *et al.* (to be published).

[41] C. Cutler and E. E. Flanagan, Phys. Rev. D **49**, 2658 (1994).

[42] Gravitational Wave OpenScience Center (GWOSC), www.gw-openscience.org/O2/.

[43] M. Vallisneri, J. Kanner, R. Williams, A. Weinstein, and B. Stephens, J. Phys. Conf. Ser. **610**, 012021 (2015).

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

[45] M. C. Miller and J. M. Miller, Phys. Rep. **548**, 1 (2015).

[46] K.-W. Lo and L.-M. Lin, Astrophys. J. **728**, 12 (2011).

[47] N. Tacik, F. Foucart, H. P. Pfeiffer, R. Haas, S. Ossokine, J. Kaplan, C. Muhlberger, M. D. Duez, L. E. Kidder *et al.*, Phys. Rev. D **92**, 124012 (2015).

[48] N. Farrow, X.-J. Zhu, and E. Thrane, Astrophys. J. **876**, 18 (2019).

[49] R. Magee, H. Fong, S. Caudill, C. Messick, K. Cannon, P. Godwin, C. Hanna, S. Kapadia, D. Meacher *et al.*, arXiv:1901.09884.

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

[51] B. Gadre, S. Mitra, and S. Dhurandhar, arXiv:1807.06803.

[52] LIGO Scientific Collaboration, LIGO Algorithm Library-LALSuite, free software (GPL), https://doi.org/10.7935/GT1W-FZ16 (2018).