Fast marginalization algorithm for optimizing gravitational wave detection, parameter estimation, and sky localization

Javier Roulet,[1,*] Jonathan Mushkin,[2] Digvijay Wadekar,[3] Tejaswi Venumadhav,[4,5] Barak Zackay,[2] and Matias Zaldarriaga[3]

[1]*TAPIR, Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, California 91125, USA*
[2]*Department of Particle Physics and Astrophysics, Weizmann Institute of Science, Rehovot 76100, Israel*
[3]*School of Natural Sciences, Institute for Advanced Study, 1 Einstein Drive, Princeton, New Jersey 08540, USA*
[4]*Department of Physics, University of California at Santa Barbara, Santa Barbara, California 93106, USA*
[5]*International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, Bangalore 560089, India*

(Received 16 April 2024; accepted 1 July 2024; published 2 August 2024)

We introduce an algorithm to marginalize the likelihood for a gravitational wave signal from a quasicircular binary merger over its extrinsic parameters, accounting for the effects of higher harmonics and spin-induced precession. The algorithm takes as input the matched-filtering time series of individual waveform harmonics against the data in all operational detectors, and the covariances of the harmonics. The outputs are the Gaussian likelihood marginalized over extrinsic parameters describing the merger time, location and orientation, along with samples from the conditional posterior of these parameters. Our algorithm exploits the waveform's known analytical dependence on extrinsic parameters to efficiently marginalize over them using a single waveform evaluation. Our current implementation achieves a 10% precision on the marginalized likelihood within ≈50 ms on a single CPU core and is publicly available through the package COGWHEEL. We discuss applications of this tool for (i) gravitational wave searches involving higher modes or precession, (ii) efficient and robust parameter estimation, and (iii) generation of sky localization maps in low latency for electromagnetic followup of gravitational-wave alerts. The inclusion of higher modes can improve the distance measurement, providing an advantage over existing low-latency localization methods.

DOI: 10.1103/PhysRevD.110.044010

I. INTRODUCTION

Gravitational wave astronomy has undergone tremendous progress over recent years, made possible by the advent of the advanced LIGO [1] and Virgo [2] detectors. In order to maximize the scientific impact of these extraordinary data, the community has developed advanced methods for identifying signals [3–12], which have yielded over a hundred detections to date [13–22]; for estimating their source parameters [23–37], which have provided invaluable insights into the astrophysics of compact binaries [38]; and for searching for short-lived electromagnetic counterparts [39–43], that enabled the identification of the kilonova from the binary neutron star merger GW170817 [44,45]. One technique that has recurrently found applications in all these fronts is the ability to marginalize the signal's likelihood over a subset of its parameters. In particular, extrinsic parameters describing the location of the observer relative to the source are the most amenable to marginalization, as their effect on the signal can be modeled analytically [39,46].

In this study, we present an algorithm for marginalizing the likelihood for a gravitational wave signal from a quasicircular binary merger over extrinsic parameters, assuming Gaussian noise and accounting for higher harmonics and spin-induced precession. The inputs to the algorithm are a set of matched-filtering time series of the waveform against the data (one time series for each harmonic mode, polarization $\{+,\times\}$ and detector) and the covariances of these components. The output is the Gaussian likelihood ratio marginalized over extrinsic parameters,

$$\bar{\mathcal{L}}(\boldsymbol{\theta}_{\rm int}) = \int \mathrm{d}\boldsymbol{\theta}_{\rm ext}\pi(\boldsymbol{\theta}_{\rm ext})\mathcal{L}(\boldsymbol{\theta}_{\rm int},\boldsymbol{\theta}_{\rm ext}), \tag{1}$$

with

*Contact author: jroulet@caltech.edu

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{p(d|\boldsymbol{\theta}, \text{signal})}{p(d|\text{Gaussian noise})}, \tag{2}$$

where d is the data, $\boldsymbol{\theta} \equiv (\boldsymbol{\theta}_{\text{int}}, \boldsymbol{\theta}_{\text{ext}})$ are the intrinsic and extrinsic parameters of the signal, and $\pi(\boldsymbol{\theta}) \equiv p(\boldsymbol{\theta}|\text{signal})$ is the prior distribution. Since the analytical dependence on $\boldsymbol{\theta}_{\text{ext}}$ is known, $\bar{\mathcal{L}}(\boldsymbol{\theta}_{\text{int}})$ can be evaluated using a single waveform query.

We envision at least three applications of this algorithm; as a piece of the detection statistic in a search incorporating higher modes or precession, as a tool for efficient and robust parameter estimation, and as a means of producing sky localization maps in low latency for electromagnetic followup of gravitational-wave alerts.

In a search, according to the Neyman-Pearson lemma, the optimal detection statistic is the likelihood ratio Λ between the two competing hypotheses; namely, that there is a signal versus only noise,

$$\Lambda(d) = \frac{p(d|\text{signal})}{p(d|\text{noise})} = \frac{\int \mathrm{d}\boldsymbol{\theta} p(d|\boldsymbol{\theta}, \text{signal}) p(\boldsymbol{\theta}|\text{signal})}{p(d|\text{noise})}. \tag{3}$$

By Eq. (2),

$$\Lambda(d) = \int \pi(\boldsymbol{\theta}) \mathcal{L}(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \cdot \frac{p(d|\text{Gaussian noise})}{p(d|\text{noise})}. \tag{4}$$

In this work we study the marginalization of the likelihood over extrinsic parameters, under the assumption of Gaussian noise in order to outline a tractable, well-defined problem. Later stages in the pipeline undertake the remaining marginalization over intrinsic parameters and apply a correction for the fact that, empirically, the noise distribution $p(d|\text{noise})$ is not Gaussian [47]. Most search pipelines have implemented techniques to perform the extrinsic-parameter marginalization, but they have generally assumed quadrupolar gravitational radiation and nonprecessing sources [3,6,17,48,49]. The contribution of this work is to include higher-order modes and spin-induced precession in the signal model, while maintaining a low computational cost compared to other components of the search. Indeed, this algorithm has been crucial in a recent search including higher modes [19].

In the context of parameter estimation, in the traditional likelihood-based paradigm a stochastic sampler is used to explore the high-dimensional parameter space, by alternatingly proposing evaluation points and computing the posterior probability density. Marginalizing the likelihood removes the extrinsic parameters from the problem, simplifying the task for the sampler. In particular, the extrinsic parameters tend to exhibit multiple modes and nonlinear degeneracies [34,50]. This approach has been pursued in the literature [46], but with implementations that either did not support higher modes and precession [51], or that were significantly more computationally intensive than the one we present here [52].

Finally, a byproduct of this algorithm is a set of extrinsic parameter samples weighted according to their conditional posterior probability, conditioned on the intrinsic parameters. If one has estimates of the intrinsic parameters (e.g., from a search pipeline), this method can be used to measure the extrinsic parameters within seconds. This mode of operation is similar to the BAYESTAR pipeline [39], except generalized to include precession and higher-order modes. This is important because higher modes are sensitive to the inclination of the binary, potentially breaking its degeneracy with the distance and localizing the source to a smaller volume [53]. Higher modes may also improve the constraints on the mass ratio of the merging objects, informing about their nature and probability of sourcing an electromagnetic counterpart.

The article is organized as follows. Section II provides a detailed description of the marginalization algorithm. Section III studies its convergence and computational cost. Section IV explores the applications to search, parameter estimation and low-latency source localization. We conclude in Sec. V. Appendix A describes various computational optimizations. Appendix B includes a code snippet demonstrating how to use our algorithm for parameter inference with the COGWHEEL software.

II. METHOD

A. Summary of the algorithm for extrinsic-parameter marginalization

Given the data and a choice of intrinsic parameters $\boldsymbol{\theta}_{\text{int}}$, we compute the marginalized likelihood Eq. (1) using a combination of integration methods; we integrate over distance by interpolating a precomputed table, over orbital phase by trapezoid quadrature, and over the remaining extrinsic parameters using adaptive importance sampling.

We first generate a large number of samples for extrinsic parameters excluding distance and orbital phase (namely; sky location, geocenter time of arrival and polarization.)[1] We draw these from a proposal distribution (described in Sec. II D) designed to be easy to compute and sample from, and to approximately match the posterior conditional on the intrinsic parameters. For each of these samples, we compute the complex inner products $(d|h^0_m)$ and $(h^0_m|h^0_{m'})$ for a signal h^0 at a fiducial distance and orbital phase, where h^0_m is the inertial-frame waveform that is generated by spherical harmonic modes with azimuthal index m in the coprecessing frame. These quantities transform in simple ways under a change of orbital phase or distance. We use the trapezoid quadrature rule to integrate over phase, and a lookup table to integrate over distance. Finally, we reweight each sample by the ratio of its posterior (marginalized over phase and distance) to the proposal probability. This yields two useful

[1]And inclination, if one restricts to aligned spins.

products: an estimate of the likelihood marginalized over all extrinsic parameters, and a set of weighted samples from the extrinsic parameter posterior. If the proposal distribution is found to inadequately describe the posterior (diagnosed as a low effective sample size) we adaptively tune the proposal and produce additional samples until we achieve convergence.

B. Waveform decomposition

In this section we write the explicit dependence of the likelihood on extrinsic parameters. We find it convenient to express the model in terms of products of various tensors, each depending on a reduced set of parameters. Throughout, we will use the subindex d to label detectors, p for polarizations $\{+,\times\}$, and (ℓ, m) for coprecessing frame harmonic modes. We will use the inner product between two time series defined as [54]

$$\langle x|y\rangle = 4\Re\int_0^\infty \mathrm{d}f\,\frac{\tilde{x}(f)\tilde{y}^*(f)}{S(f)}, \tag{5}$$

where S is the one-sided noise power spectrum, and a summation over detectors is assumed.

We start with the standard Gaussian likelihood ratio $\mathcal{L}$ (we henceforth refer to $\mathcal{L}$ simply as the likelihood),

$$\ln\mathcal{L}(\boldsymbol{\theta}) = \langle d|h(\boldsymbol{\theta})\rangle - \frac{1}{2}\langle h(\boldsymbol{\theta})|h(\boldsymbol{\theta})\rangle. \tag{6}$$

Here, d is the strain data and h the model waveform. Extrinsic parameters modify the waveform in a well-understood way,[2]

$$\begin{aligned}&\tilde{h}_d(f;\boldsymbol{\theta}_{\rm int},\psi,\hat{\boldsymbol{n}},t_\oplus,\phi_{\rm ref},D)\\ &=\sum_{m=1}^{\ell_{\max}}\sum_{p\in\{+,\times\}}\tilde{h}_{mp}(f;\boldsymbol{\theta}_{\rm int})\frac{F_{dp}(\hat{\boldsymbol{n}},\psi)\mathrm{e}^{-\mathrm{i}2\pi f t_d(t_\oplus,\hat{\boldsymbol{n}})}\mathrm{e}^{\mathrm{i}m\phi_{\rm ref}}}{D},\end{aligned} \tag{7}$$

where we have grouped the harmonics by m,

$$\tilde{h}_{mp}(f;\boldsymbol{\theta}_{\rm int}) \coloneqq \sum_\ell \tilde{h}_{\ell mp}(f;\boldsymbol{\theta}_{\rm int}, D=1,\phi_{\rm ref}=0). \tag{8}$$

Here, the indices ℓ, m denote spherical-harmonic modes in the co-precessing frame, but the harmonic $\tilde{h}_{\ell mp}$ is the inertial-frame waveform generated by the "twisting up" procedure operating on this coprecessing harmonic [58]. In particular, the $\tilde{h}_{\ell mp}$ has different spherical harmonic content in the inertial frame.

[2]We follow the default LALSimulation convention and use labels $m>0$, understanding that the m and $-m$ coprecessing harmonics are summed together using $\tilde{h}_{\ell m}(f) = (-1)^\ell \tilde{h}_{\ell,-m}(-f)$, with $f>0$ [55–57].

Reading Eq. (7) from the left, the right-hand side is interpreted as follows. The source emits polarized waves h_{mp}, to which the detector has an antenna response $F_{dp}(\hat{\boldsymbol{n}},\psi)$ that depends on the geometrical configuration. The signal arrives at each detector at time

$$t_d(t_\oplus,\hat{\boldsymbol{n}}) = t_\oplus - \boldsymbol{r}_d\cdot\hat{\boldsymbol{n}}/c, \tag{9}$$

which is the overall time of arrival at geocenter $t_\oplus$ plus a time-of-travel correction that depends on the location $\boldsymbol{r}_d$ of the detector projected onto the line of sight $\hat{\boldsymbol{n}}$. Each coprecessing harmonic $\tilde{h}_{\ell m}$ transforms according to $\mathrm{e}^{\mathrm{i}m\phi_{\rm ref}}$ under a rotation in the plane of the binary.[3] Lastly, the waveform amplitude decays in inverse proportion to the luminosity distance to the source, D.

For precessing signals, the inclination of the orbit is frequency-dependent, and therefore we will treat it as an intrinsic (nonmarginalized) parameter. For nonprecessing (aligned-spin) systems, the inclination can be treated analytically by replacing $\mathrm{e}^{\mathrm{i}m\phi_{\rm ref}}$ by the spin-weighted harmonic ${}_{-2}Y_{\ell m}(\iota,\phi_{\rm ref})$ in Eq. (7) [59].

Using Eq. (7), we can rewrite Eq. (6) in terms of factors that depend separately on the intrinsic or the extrinsic parameters,

$$\langle d|h\rangle = \frac{1}{D}\Re\left\{\sum_m \mathrm{e}^{-\mathrm{i}m\phi_{\rm ref}}\sum_{d,p}F_{dp}(\hat{\boldsymbol{n}},\psi)z_{mpd}(t_d(t_\oplus,\hat{\boldsymbol{n}});\boldsymbol{\theta}_{\rm int})\right\}, \tag{10}$$

where the time series

$$z_{mpd}(t;\boldsymbol{\theta}_{\rm int}) \coloneqq 4\int_0^\infty \mathrm{d}f\,\frac{\tilde{d}_d(f)\tilde{h}^*_{mp}(f;\boldsymbol{\theta}_{\rm int})}{S_d(f)}\mathrm{e}^{\mathrm{i}2\pi f t} \tag{11}$$

is the complex matched-filter output [60] of the waveform's mode m and polarization p in the dth detector. In practice only a short interval of time around the peak is needed. Similarly,

$$\begin{aligned}\langle h|h\rangle = &\frac{1}{D^2}\sum_{m,m'}\mathrm{e}^{\mathrm{i}(m'-m)\phi_{\rm ref}}\\ &\times\sum_{d,p,p'}c_{mm'pp'd}(\boldsymbol{\theta}_{\rm int})F_{dp}(\hat{\boldsymbol{n}},\psi)F_{dp'}(\hat{\boldsymbol{n}},\psi)\end{aligned} \tag{12}$$

with

$$c_{mm'pp'd}(\boldsymbol{\theta}_{\rm int}) = 4\int_0^\infty \mathrm{d}f\,\frac{\tilde{h}_{mp}(f;\boldsymbol{\theta}_{\rm int})\tilde{h}^*_{m'p'}(f;\boldsymbol{\theta}_{\rm int})}{S_d(f)}. \tag{13}$$

[3]When we vary the orbital phase $\phi_{\rm ref}$ we hold the black hole spins fixed with respect to the orbital angular momentum and the direction of propagation (not with respect to the orbital separation vector) [34].

Substituting Eqs. (10) and (12) into (6), we have decomposed the likelihood into factors that depend either on extrinsic or intrinsic parameters. The matched-filter time series $z_{mpd}(t;\boldsymbol{\theta}_{\rm int})$ and covariances $c_{mm'pp'd}(\boldsymbol{\theta}_{\rm int})$ encapsulate all the dependence on intrinsic parameters and will be the inputs to our computation.

C. Phase and distance integration

The distance, orbital phase and polarization are simpler to marginalize than other extrinsic parameters, because they do not affect the times of arrival, and hence their effect on the waveform is independent of frequency. This makes it inexpensive to vary these parameters, as the data can first be compressed to a few numbers with the frequency axis collapsed. Here we marginalize the distance and orbital phase explicitly, and in Sec. II D we integrate the remaining extrinsic parameters using importance sampling. Since the orbital phase and polarization are largely degenerate [23,34], it suffices to marginalize only one of the two at high resolution; we will do the phase.

Holding all other parameters fixed, the distance-marginalized likelihood $\bar{\mathcal{L}}_D$ can be expressed in terms of only two values, namely the inner products $\langle d|h_1\rangle$ and $\langle h_1|h_1\rangle$ for a waveform at unit distance, $h_1 \coloneqq h(D=1)$,

$$\bar{\mathcal{L}}_D(\langle d|h_1\rangle, \langle h_1|h_1\rangle) = \int \mathrm{d}D\, \pi(D) \exp\left(\frac{\langle d|h_1\rangle}{D} - \frac{\langle h_1|h_1\rangle}{2D^2}\right). \tag{14}$$

Following Singer and Price [39], after suitable rescaling and reparametrization Eq. (14) can be efficiently evaluated by 2D interpolation of a precomputed lookup table. This is possible since neither higher modes nor precession modify the dependence of the waveform on distance.

However, higher modes do change the dependence on orbital phase, and hence we cannot marginalize the phase analytically, as usually done for quadrupolar waveforms. Instead, we use trapezoid quadrature, which performs adequately since the likelihood is a periodic function of the orbital phase. To integrate Eq. (14), it suffices to evaluate $\langle d|h_1\rangle$ and $\langle h_1|h_1\rangle$ on a regular grid $\{\phi_{{\rm ref},o}\}$ covering the orbital phases, where the subindex o runs through $1,\ldots,N_\phi$. For $\langle d|h_1\rangle$, we compute

$$\langle d|h_1\rangle_o = \Re \sum_m (d|h_1)_m \Phi_{mo}, \tag{15}$$

where

$$(d|h_1)_m \coloneqq \sum_{p,d} F_{dp}(\hat{\boldsymbol{n}},\psi) z_{mpd}(t_d(t_\oplus,\hat{\boldsymbol{n}});\boldsymbol{\theta}_{\rm int}) \tag{16}$$

is obtained by cubic spline interpolation of the time series z_{mpd} and

$$\Phi_{mo} \coloneqq \exp(\mathrm{i}\, m\, \phi_{{\rm ref},o}) \tag{17}$$

is precomputed. For $\langle h_1|h_1\rangle$, similarly

$$\langle h_1|h_1\rangle_o = \Re \sum_{m,m'} (h_1|h_1)_{mm'} \Phi_{mm'o}, \tag{18}$$

$$(h_1|h_1)_{mm'} \coloneqq \sum_{d,p,p'} c_{mm'pp'd}(\boldsymbol{\theta}_{\rm int}) F_{dp}(\hat{\boldsymbol{n}},\psi) F_{dp'}(\hat{\boldsymbol{n}},\psi) \tag{19}$$

$$\Phi_{mm'o} \coloneqq \exp[\mathrm{i}(m'-m)\phi_{{\rm ref},o}]. \tag{20}$$

Using Eqs. (14), (15), and (18) we obtain the likelihood marginalized over orbital phase and distance,

$$\bar{\mathcal{L}}_{\phi D}(\psi,\hat{\boldsymbol{n}},t_\oplus;\boldsymbol{\theta}_{\rm int}) \approx \frac{1}{N_\phi} \sum_{o=1}^{N_\phi} \bar{\mathcal{L}}_D(\langle d|h\rangle_o, \langle h|h\rangle_o). \tag{21}$$

From these data products, posterior samples of distance and orbital phase can also be readily generated. Phase samples can be drawn from the grid according to the weights [the summands in Eq. (21)] and then distance samples can be produced from the integrand of Eq. (14) with inverse transform sampling.

D. Time, sky location, and polarization integration

We perform the integral in Eq. (1) over the remaining extrinsic parameters (sky location $\hat{\boldsymbol{n}}$, time of arrival $t_\oplus$, and polarization angle ψ) using importance sampling [61,62]. We will choose a proposal distribution $p(\psi,t_\oplus,\hat{\boldsymbol{n}})$, and construct it in a way that will allow us to easily generate samples from it. The marginal likelihood will be estimated from those samples as

$$\bar{\mathcal{L}}(\boldsymbol{\theta}_{\rm int}) \approx \frac{1}{N} \sum_{i=1}^{N} \frac{\pi(\psi^i,t^i_\oplus,\hat{\boldsymbol{n}}^i)\bar{\mathcal{L}}_{\phi D}(\psi^i,\hat{\boldsymbol{n}}^i,t^i_\oplus;\boldsymbol{\theta}_{\rm int})}{p(\psi^i,t^i_\oplus,\hat{\boldsymbol{n}}^i)}. \tag{22}$$

The weighted samples also allow to sample the conditional posterior $p(\boldsymbol{\theta}_{\rm ext}|d,\boldsymbol{\theta}_{\rm int})$, by simply drawing according to the weights [i.e., the summands in Eq. (22)].

The variance of the estimator in Eq. (22) is highly sensitive to the choice of proposal: it vanishes when p is proportional to the integrand [i.e., the conditional posterior for $(\psi,t_\oplus,\hat{\boldsymbol{n}})$ in the numerator of Eq. (22)], on the other hand, it diverges if p has a tighter support. When this happens, a small number of samples in the tail of p are disproportionately upweighted and dominate the sum. Thus, we will design the proposal to approximately match the conditional posterior, erring on the side of having heavier tails. Having reduced the dimensionality of the quasi-Monte Carlo integral Eq. (31) by explicitly integrating out the orbital phase and distance (Sec. II C) improves its efficiency, especially considering that the orbital phase is very well-measured when other parameters are kept fixed.

For ψ, we simply use its uniform prior as proposal, based on the heuristic that it is rarely well constrained due to degeneracy with the orbital phase $\phi_{\rm ref}$ [23,34].

In contrast, $\hat{\boldsymbol{n}}$ and $t_\oplus$ are usually measured very well compared to the size of their prior, calling for a more sophisticated proposal. The constraints on these parameters are largely driven by the measurement of the arrival times at the individual detectors. Notably, it is possible to estimate these arrival times separately at each detector, and, furthermore, to precompute their relation to $t_\oplus$ and $\hat{\boldsymbol{n}}$ independently of the data. With these insights, we follow [17,51] and specify our proposal distribution over $(t_\oplus, \hat{\boldsymbol{n}})$ with the help of an auxiliary proposal $P(\boldsymbol{\tau})$ for the discretized times of arrival τ_d at each detector.[4]

1. Proposal for arrival direction and geocenter time

We choose a timescale Δ sufficiently small to resolve the autocorrelation length of the whitened template (thus, structure in the matched-filtering time series), and discretize the time axis at this resolution. We then partition the $(t_\oplus, \hat{\boldsymbol{n}})$ space into exhaustive disjoint regions $\mathcal{D}(\boldsymbol{\tau})$, where $\boldsymbol{\tau} \equiv \{\tau_d\}$ defines a discrete time of arrival at each detector, and $\mathcal{D}(\boldsymbol{\tau})$ is the domain of arrival time and sky location consistent with those $\boldsymbol{\tau}$. Our criterion for consistency is that the time of arrival at the first detector, and the time delays between the first detector and the others, match those of $\boldsymbol{\tau}$ to a precision $\Delta/2$,

$$\mathcal{D}(\boldsymbol{\tau}) = \left\{ t_\oplus, \hat{\boldsymbol{n}} : |t_{d_0}(t_\oplus, \hat{\boldsymbol{n}}) - \tau_{d_0}| < \frac{\Delta}{2} \right. \\ \left. \wedge\, |\delta t_d(\hat{\boldsymbol{n}}) - \delta\tau_d| < \frac{\Delta}{2} \right\}, \tag{23}$$

with

$$\delta t_d(\hat{\boldsymbol{n}}) \coloneqq t_d(t_\oplus, \hat{\boldsymbol{n}}) - t_{d_0}(t_\oplus, \hat{\boldsymbol{n}}) \tag{24}$$

$$\delta\tau_d \coloneqq \tau_d - \tau_{d_0}, \tag{25}$$

where d_0 is the arbitrary first detector. The time delays $\boldsymbol{\delta t}$ and $\boldsymbol{\delta\tau}$ have $N_{\rm detectors} - 1$ components each, and $\boldsymbol{\delta t}$ is independent of $t_\oplus$.

Our strategy is to first draw samples $\boldsymbol{\tau}^i$ from a proposal $P(\boldsymbol{\tau})$ (described later), and to each assign a $t_\oplus^i, \hat{\boldsymbol{n}}^i \sim \pi(t_\oplus, \hat{\boldsymbol{n}}|\boldsymbol{\tau}^i)$ drawn from the restricted prior, by means of a precomputed mapping that we construct as follows.

Ahead of time, we draw a large number of samples (10^6 is our current default) isotropically distributed in the sky, in terms of Earth-fixed coordinates (latitude and longitude). For each sample we compute $\boldsymbol{\delta t}(\hat{\boldsymbol{n}})$, and based on this we assign it to the nearest discretized time-delay $\boldsymbol{\delta\tau}$. The resulting map is shown in Fig. 1 for the example case of a Hanford-Livingston-Virgo network. Given a $\boldsymbol{\delta\tau}$, the mapping provides a set of consistent sky location samples [color coded in Fig. 1(a)]. Once a detector time sample $\boldsymbol{\tau}^i$

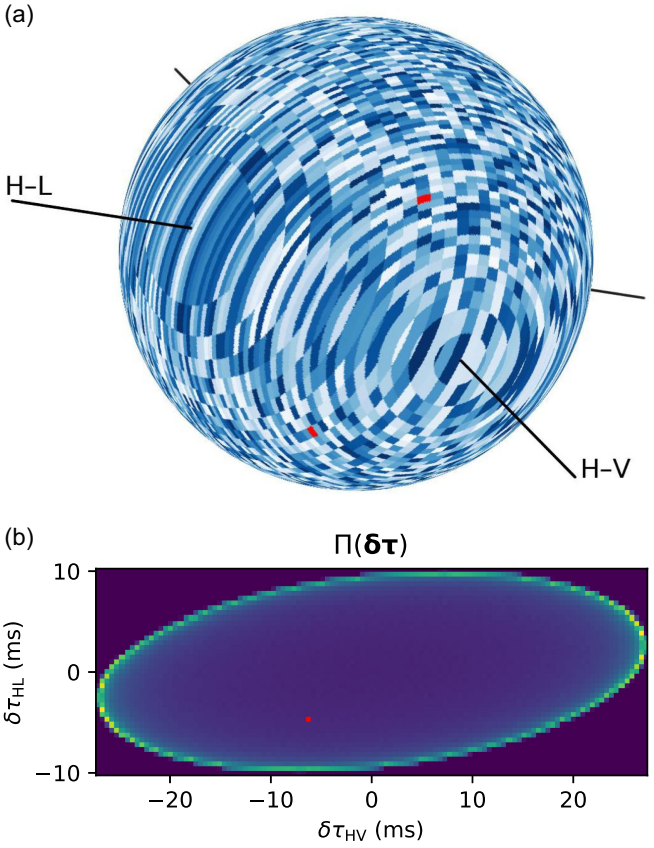

FIG. 1. Partition of the space of arrival directions $\hat{\boldsymbol{n}}$ by discretized time delays between detectors $\boldsymbol{\delta\tau}$, for the particular case of a Hanford-Livingston-Virgo network. We use this mapping to efficiently assign a consistent arrival direction to a proposed set of discrete arrival times at the detectors, during the importance-sampling marginalization over $\hat{\boldsymbol{n}}$. (a) The time delay in each pair of detectors defines a ring in the sky, perpendicular to the corresponding detector separation vector (black axis). For the 3-detector network shown here, each subregion is the intersection of two rings. For a 2-detector network, the subregions would instead be annular, and for a single detector there would be one region covering the entire sky [For upcoming detector networks consisting of more than three detectors, the strategy of proposing arrival times at each detector would result in overconstrained sky locations. The method can be generalized by using only the most sensitive three detectors for the proposal (and all the detectors for reweighting) in that case)]. (b) Prior for discretized time delays between detectors, proportional to the solid angle of the associated patch of the sky. One particular $\boldsymbol{\delta\tau}$ is highlighted in red in both plots. Two sky patches, symmetric about the plane containing the detectors, share the same delays. The resolution of the map was lowered for visual clarity; by default we use a 4× higher one of 8192 Hz. For a 2-detector network the prior would be a one-dimensional array, and for a single detector it would be a scalar number.

[4]We use capital letters for discrete distributions, lowercase for continuous distributions.

has been proposed, we compute $\boldsymbol{\delta\tau}^i$ and assign a sky location sample $\hat{\boldsymbol{n}}^i$ from the corresponding entry of the mapping. We also draw a time of arrival at the first detector $t^i_{d_0}$ uniformly within $\tau^i_{d_0} \pm \Delta/2$, and solve Eq. (9) to obtain $t^i_\oplus$.

As a useful byproduct, this map allows us to obtain the physical prior on the time-delays $\Pi(\boldsymbol{\delta\tau})$, estimated as the fraction of samples assigned to each $\boldsymbol{\delta\tau}$. We show this prior in Fig. 1(b), it will be necessary later. In particular, unphysical $\boldsymbol{\delta\tau}$ (say, delays longer than the gravitational wave travel time between detectors) get no sky location samples assigned. To keep the variance of this estimate low, we use a quasirandom Halton sequence to draw the sky locations, which covers the sky more uniformly than e.g., random sampling or a spherical grid.

Note that near the plane containing the detectors, perpendicular displacements produce only quadratic shifts in the arrival times. Accordingly, in Fig. 1(a) these cells are elongated and cover a large solid angle, and have a large prior $\Pi(\boldsymbol{\delta\tau})$ in Fig. 1(b). We expect that sources near the plane of the detectors will have a relatively poor sky location measurement perpendicular to the plane.

The proposal density that results from the above process is

$$\begin{aligned} p(t^i_\oplus, \hat{\boldsymbol{n}}^i) &= \sum_{\boldsymbol{\tau}} P(\boldsymbol{\tau}) \pi(t^i_\oplus, \hat{\boldsymbol{n}}^i | \boldsymbol{\tau}) \\ &= P(\boldsymbol{\tau}^i) \pi(t^i_\oplus, \hat{\boldsymbol{n}}^i | \boldsymbol{\tau}^i). \end{aligned} \tag{26}$$

The second line follows because the restricted prior is zero for detector arrival times that are inconsistent with the sample: $\pi(t^i_\oplus, \hat{\boldsymbol{n}}^i | \boldsymbol{\tau} \neq \boldsymbol{\tau}^i) = 0$. Combining Eqs. (22) and (26), we obtain

$$\bar{\mathcal{L}}(\boldsymbol{\theta}_{\rm int}) \approx \frac{1}{N} \sum_{i=1}^{N} \frac{\pi(t^i_\oplus, \hat{\boldsymbol{n}}^i) \bar{\mathcal{L}}_{\phi D}(\psi^i, \hat{\boldsymbol{n}}^i, t^i_\oplus; \boldsymbol{\theta}_{\rm int})}{P(\boldsymbol{\tau}^i) \pi(t^i_\oplus, \hat{\boldsymbol{n}}^i | \boldsymbol{\tau}^i)}. \tag{27}$$

The prior in the numerator of Eq. (27) is separable and uniform, $\pi(t_\oplus)\pi(\hat{\boldsymbol{n}}) = \text{const}$. Since there is no natural domain for the time, we will adopt a dimensionless prior $\pi(t_\oplus) = 1$ and recognize that the marginalized likelihood has units of time. The restricted prior $\pi(t_\oplus, \hat{\boldsymbol{n}} | \boldsymbol{\tau})$ is proportional to $\pi(t_\oplus, \hat{\boldsymbol{n}})$ but integrates to 1 over $\mathcal{D}(\boldsymbol{\tau})$, hence, their ratio in Eq. (27) is

$$\begin{aligned} \frac{\pi(t_\oplus, \hat{\boldsymbol{n}})}{\pi(t_\oplus, \hat{\boldsymbol{n}} | \boldsymbol{\tau})} &\equiv \Pi(\boldsymbol{\tau}) \\ &= \int_{\mathcal{D}(\boldsymbol{\tau})} \mathrm{d}t_\oplus \mathrm{d}\hat{\boldsymbol{n}} \pi(t_\oplus) \pi(\hat{\boldsymbol{n}}) \\ &= \Delta \cdot \Pi(\boldsymbol{\delta\tau}). \end{aligned} \tag{28}$$

In Eq. (28), the $t_\oplus$ integral equals Δ, and the $\hat{\boldsymbol{n}}$ integral yields $\Pi(\boldsymbol{\delta\tau})$, i.e., the fraction of the sky compatible with the time delays that we introduced in Fig. 1(b).

Substituting Eq. (28) in (27), we arrive at

$$\bar{\mathcal{L}}(\boldsymbol{\theta}_{\rm int}) \approx \frac{1}{N} \sum_{i=1}^{N} \frac{\Delta \cdot \Pi(\boldsymbol{\delta\tau}^i)}{P(\boldsymbol{\tau}^i)} \bar{\mathcal{L}}_{\phi D}(\psi^i, \hat{\boldsymbol{n}}^i, t^i_\oplus; \boldsymbol{\theta}_{\rm int}). \tag{29}$$

2. Adaptive multiple importance sampling

As discussed above, the variance of the importance sampling integral can be large if the proposal is misspecified. The most sensitive component of the proposal is the auxiliary distribution of discrete detector arrival times $P(\boldsymbol{\tau})$, as it is responsible for the largest reduction in phase space volume. To make $P(\boldsymbol{\tau})$ robust to an eventual initial misestimation, we will allow the option of adapting it as needed by iteratively proposing distributions $P^{(j)}$, that attempt to cover any problematic regions where the previous proposals were too narrow. The total proposal is a mixture of the form,

$$\begin{aligned} P(\boldsymbol{\tau}) &= \sum_j \alpha_j P^{(j)}(\boldsymbol{\tau}) \\ &= \sum_j \frac{N_j}{N} \prod_d P^{(j)}_d(\tau_d), \end{aligned} \tag{30}$$

with $\sum_\tau P^{(j)}_d(\tau) = 1$. That is, we define a series of adaptive proposals $P^{(j)}(\boldsymbol{\tau})$, each factorizable over detectors.[5] This property makes drawing samples of $\boldsymbol{\tau} \sim P^{(j)}$ a simple task, as the τ_d are drawn independently from one-dimensional distributions. This task can be achieved with the inverse transform sampling technique, which is efficient and furthermore facilitates the use of quasi-Monte Carlo integration, as we will explain in Sec. II D 5. Every time we add a new proposal $P^{(j)}$, we draw N_j samples from it, and combine them with the previous ones using the so-called balance heuristic $\alpha_j = N_j/N$, with $N = \sum_j N_j$. Altogether, Eqs. (29) and (30) become

$$\bar{\mathcal{L}}(\boldsymbol{\theta}_{\rm int}) \approx \sum_{i=1}^{N} w_i, \tag{31}$$

$$w_i = \frac{\Delta \cdot \Pi(\boldsymbol{\delta\tau}^i)}{\sum_j (N_j \prod_d P^{(j)}_d({\tau_d}^i))} \bar{\mathcal{L}}_{\phi D}(\psi^i, \hat{\boldsymbol{n}}^i, t^i_\oplus; \boldsymbol{\theta}_{\rm int}). \tag{32}$$

The importance sampling weights w_i can be used to estimate the effective number of samples

[5]Note that the total proposal $P(\boldsymbol{\tau})$ is not factorizable.

$$N_{\rm eff} \equiv \frac{(\sum_i w_i)^2}{\sum_i w_i^2}. \tag{33}$$

We set a threshold $N_{\rm eff}^{\rm min}$, and iteratively add proposals $P^{(j)}$ until the effective sample size meets this threshold (or a maximum number of proposals is reached). Every time a new $P^{(j)}$ is added, we draw N_j samples from it, update the weights of all samples and recompute $N_{\rm eff}$. In Sec. III we will confirm that $N_{\rm eff}$ is a good tracer of the precision of the importance sampling integral.

3. Initial proposal distribution of arrival times at detectors

We choose the initial ($j = 0$) proposal distribution of detector arrival times for each detector d based on the matched-filtering time series and covariances. We adopt the following functional form:

$$P_d^{(0)}(\tau) = \Pi_d(\tau) \exp[\beta_d \ln \hat{\mathcal{L}}_d(\tau)], \tag{34}$$

where $\hat{\mathcal{L}}_d$ is an approximate likelihood, $0 < \beta_d \leq 1$ is a tempering factor and Π_d is a prior. For the likelihood we use

$$\ln \hat{\mathcal{L}}_d(\tau) = \frac{(\sum_{m,p} |z_{mpd}(\tau)|)^2}{2\sum_{m,p} c_{mmppd}}. \tag{35}$$

This expression follows from approximating that different modes and polarizations are orthogonal and have independent phases, and then maximizing Eq. (6) over these phases and the distance.

The tempering factors β_d are intended to make the initial proposal broader, and therefore more robust against missing the support of the posterior. How to choose them depends on the application, for example in a search they may be fixed heuristically (e.g., $\beta = 0.5$), or in parameter estimation they can be tuned at the beginning to maximize $N_{\rm eff}$ for a reference waveform.

Finally, we use the priors Π_d to incorporate information about the physically allowed time delays between detectors: if the signal is weak in a detector, the arrival time at that detector might be meaningfully constrained by data in other detectors. While the true prior on arrival times τ_d is correlated among detectors, in order to draw arrival time samples easily we require that the proposal distributions $P_d^{(j)}$ are uncorrelated [see Eq. (30)]. We circumvent this by conditioning the proposal in one detector on the other detectors' proposal distributions rather than on the individual values of the arrival time samples. We achieve this as follows. Once we have computed the likelihood (35), we sort the detectors by decreasing $\max_\tau \hat{\mathcal{L}}_d$. In the first (loudest) detector $d = 1$, where the likelihood best constrains the time of arrival, we use a uniform prior

$$\Pi_1(\tau) = \text{const}, \tag{36}$$

as this defines $P_1^{(0)}(\tau)$. For the second detector we condition the prior on our knowledge of $P_1^{(0)}$,

$$\Pi_2 = P_1^{(0)} * \Pi_{21}, \tag{37}$$

i.e., we use the proposal for the time of arrival at the first detector convolved with the prior distribution of time delays $\tau_{21} = \tau_2 - \tau_1$ to the second detector. This incorporates the information that there is a maximum allowed time delay. We compute the prior for the arrival time at the third detector in a conceptually similar way, where now Π_3 is informed by $P_1^{(0)}$ and $P_2^{(0)}$. We first estimate the time delay between the first two detectors by cross-correlating their proposals, $P(\tau_{21}) = P_1^{(0)} \star P_2^{(0)}$. We marginalize over this distribution to obtain $\Pi(\tau_{31}|P(\tau_{21})) = \sum_{\tau_{21}} \Pi(\tau_{21}, \tau_{31}) P(\tau_{21})$, and finally arrive at

$$\Pi_3 = P_1^{(0)} * \Pi_{31|21}. \tag{38}$$

4. Adaptation

After iteration J of the adaptation, we have the set of samples $\{\boldsymbol{\tau}^i\}, i = \{1, \ldots, \sum_j^J N_j\}$ proposed so far, along with their weights w_i.

For the following proposal $P_d^{(J+1)}(\tau_d)$ we aim to match the proposal to the posterior, perhaps with heavier tails. We obtain a measurement of the detector arrival time posterior by kernel density estimation (KDE) on the existing samples: we construct a histogram of $\{\tau_d^i\}$ weighted by w_i and convolve it with a suitably chosen kernel. We use a heavy-tailed Cauchy kernel $K(\delta\tau; \Sigma) \propto (\delta\tau^2 + \Sigma^2)^{-1}$. We set the kernel width Σ in each detector using Silverman's rule of thumb [63] with a lower bound Δ: $\Sigma_d = \max\{\Delta, (4N_{\rm eff}/3)^{-1/5}\sigma_d\}$, where σ_d is the weighted standard deviation of the sample of $\{\tau_d^i\}$.

As a new proposal we use a hybrid between this KDE and the previous proposal,

$$P_d^{(J+1)}(\tau_d) = \frac{1}{2}\left(\text{KDE}(\tau_d) + \sum_j^J P_d^{(j)}(\tau_d)\right). \tag{39}$$

This increases the stability of the adaptation and handles satisfactorily the generic situation in which the original proposal was adequate in some detectors and not others.

This adaptation step is illustrated in the second and third rows of Fig. 2.

5. Quasi-Monte Carlo

In order to further reduce the variance of the $(\psi, \hat{\boldsymbol{n}}, t_\oplus)$ integral estimated in Eq. (31), we jointly draw the samples of detector arrival times, polarization and subgrid time shift using quasi-Monte Carlo [61]. By design, the proposal is

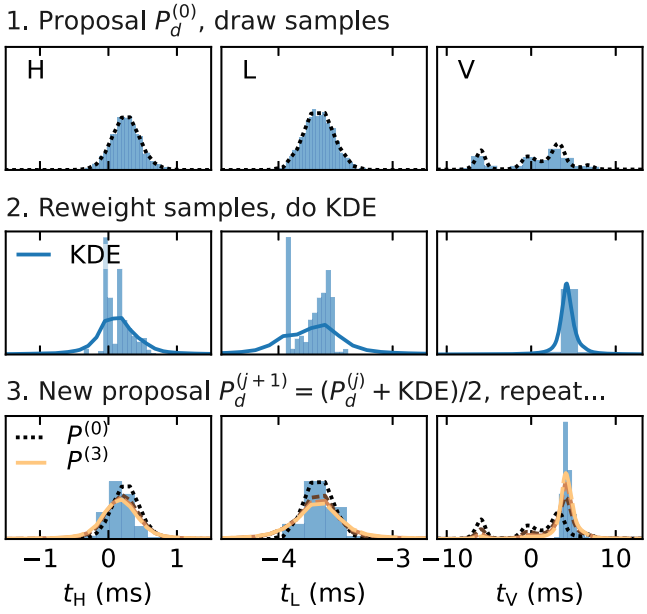

FIG. 2. To reduce the variance of the importance sampling marginalization, we use an adaptive proposal distribution for the detector arrival times. Top: an initial proposal is generated in each detector from the matched-filtering time series (Sec. II D 3). Sets of discrete arrival times $\boldsymbol{\tau}$ are sampled from these with quasi-Monte Carlo. Physical parameters are assigned to each set of detector arrival times (Sec. II D 1). Center: the physical samples are reweighted according to the ratio of their (coherent) posterior to the proposal [Eq. (32)], and used to estimate the probability density in each detector via KDE (Sec. II D 4). Bottom: the proposal is updated by averaging it with the KDE. Previous samples are kept, and their weights updated to reflect the change of the proposal. The process is repeated until the effective number of samples is satisfactory.

factorizable in all these variables. In each dimension we use inverse transform sampling, i.e., use the cumulative of the proposal, u, as coordinate, so that the proposal becomes the uniform distribution,

$$u_a(x_a) \coloneqq \int_{-\infty}^{x_a} p_a(x'_a)\mathrm{d}x'_a,$$
$$\Rightarrow u_a \sim \mathrm{U}(0,1), \tag{40}$$

where a labels each dimension (arrival time at each detector, subgrid timeshift, polarization) and p_a is the corresponding proposal. Instead of drawing the $\{\boldsymbol{u}^i\}$ independently, we select them according to a scrambled Halton sequence. This introduces correlations between the samples, that decrease the variance of the estimator Eq. (31) by making the average covariance of the weights negative. We then invert Eq. (40) to obtain the physical quantities $\{\psi^i, \boldsymbol{\tau}^i, t^i_{d_0}\}$.

III. CONVERGENCE AND PERFORMANCE

In Fig. 3 we study the accuracy of the marginalization algorithm by comparing multiple estimates to a high resolution result that serves as ground truth. We find that for different events, intrinsic parameter values, realization of importance samples, and number of optimizations, the effective number of samples remains a good tracer of the error in the computation. This is important because the effective number of samples can be computed from the available importance weights at negligible cost. In the figure legend we report the maximum-likelihood fit of a model in which the $\ln \bar{\mathcal{L}}$ errors are normally distributed with a variance that follows a power-law on the effective number of samples. Notably, very precise estimates of the marginal likelihood can be obtained as needed by increasing the number of samples.

In Fig. 4 we show the computational cost of the extrinsic-parameter marginalization using our implementation of the algorithm. The number of effective samples increases linearly or faster with the computational effort; the latter situation is indicative of cases where the proposal adaptation makes an impact. We typically achieve 10% precision within 50 ms with a sizable variance contingent on the event and intrinsic parameter values.

In both Figs. 3 and 4, the intrinsic-parameter evaluation points were chosen at random from the chain of proposals made by the NAUTILUS sampler [64] in a parameter estimation run on data with a synthetic signal injected (see Sec. IV B for additional details), to ensure that they are representative of real-world applications.

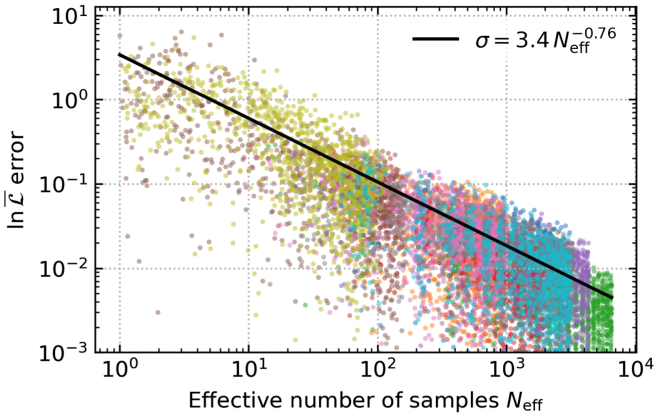

FIG. 3. Convergence test of our marginalization algorithm. We show the importance sampling error in the marginalized likelihood $\bar{\mathcal{L}}(d|\boldsymbol{\theta}_{\mathrm{int}})$ versus effective number of samples in the estimator [Eq. (33)]. Each point represents a single marginalized likelihood estimate. Different colors correspond to different events d and intrinsic-parameter values $\boldsymbol{\theta}_{\mathrm{int}}$. Within each color, points differ in the number of proposal adaptations performed and the realization of extrinsic-parameter importance samples (Sec. II D). A single fit to the errors is found to describe all examples reasonably well.

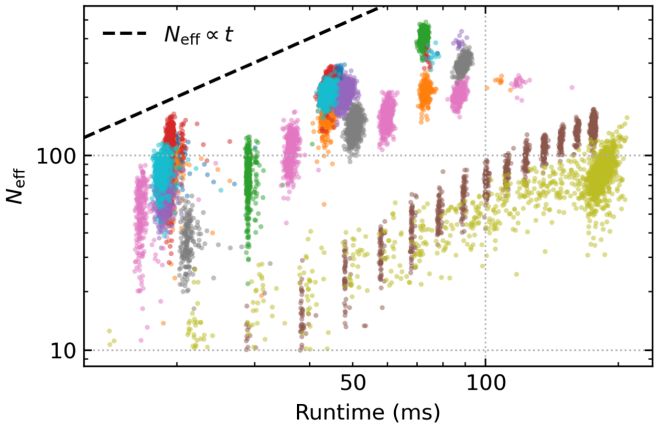

FIG. 4. Computational cost of the marginalization over extrinsic parameters. We plot the effective number of samples in the importance sampling estimator as a function of the time spent by one CPU core. As in Fig. 3, different colors correspond to different events d and intrinsic-parameter values $\boldsymbol{\theta}_{\rm int}$. While the computational cost depends considerably on the particular event and parameter values, in most cases $N_{\rm eff} \gtrsim 100$ ($\lesssim$10% uncertainty in $\bar{\mathcal{L}}$) is achieved within 50 ms.

IV. APPLICATIONS

A. Search for binary mergers

As argued in Sec. I, optimally searching for gravitational wave events requires ranking candidates by their likelihood marginalized over the nuisance parameters of the signal model. The method introduced in Sec. II allows us to marginalize over the extrinsic parameters accounting for higher-order modes, enabling searches sensitive to this type of signals. Indeed, we have used this algorithm in a search pipeline that incorporates higher modes [19].

In terms of sensitivity, one could compare this method to a common alternative to marginalization, which is to maximize the likelihood over nuisance parameters. That approach is reasonably close to optimal in the case where there are a few well-measured parameters. However, higher-order modes introduce nuanced details in the waveform shape that greatly increase the diversity of waveforms, while at the same time they are a subdominant perturbation for most parameter values. The diversity of waveforms incurs a large trials factor, to the point where the inclusion of higher modes may in fact degrade the sensitivity of a maximum-likelihood search [65]. This degradation happens because the likelihood may be maximized for fine-tuned configurations that are not representative of the generic solutions, whereas marginalization correctly penalizes such configurations. For example, oftentimes the likelihood is maximized for a nearly edge-on inclination, but this solution is penalized when we marginalize over distance, as highly inclined systems are observable in a smaller volume. From first principles, with the marginalization statistic the more accurate model (including higher harmonics) is guaranteed to have a better sensitivity by the Neyman-Pearson lemma.

In terms of computational cost, the most expensive inputs to the algorithm are the matched-filtering time series $z_{mpd}(t;\boldsymbol{\theta}_{\rm int})$ for each of the harmonic modes. For a search, these are computed by means of fast Fourier transforms. This requires constructing the template bank in terms of the individual harmonics, which turns out to be convenient since (in line with the above discussion) it leads to banks of a $\sim$100 times smaller size compared to banks of fully specified templates [66]. The covariances $c_{mm'pp'd}(\boldsymbol{\theta}_{\rm int})$ are time-independent (except for slow variations due to the nonstationarity of the noise) and therefore inexpensive. The efficiency of our marginalization routine enabled us to use it as component of the ranking score on $\sim 10^7$ foreground and background (i.e., with artificial time shifts between detectors applied [4,8]) triggers in a recent search including higher-order modes [19], without it becoming a computational bottleneck.

While the likelihood ratio in this work was derived under the assumption of Gaussian noise, a later stage in the pipeline applies a correction for the fact that the empirical noise distribution is not Gaussian [47].

B. Parameter estimation

In this section we demonstrate the applicability of the extrinsic parameter marginalization to parameter estimation. We use a general purpose stochastic sampler to explore the intrinsic-parameter posterior,

$$p(\boldsymbol{\theta}_{\rm int}|d) = \pi(\boldsymbol{\theta}_{\rm int})\bar{\mathcal{L}}(\boldsymbol{\theta}_{\rm int}). \tag{41}$$

To reconstruct the full distribution, for each intrinsic-parameter sample, we select extrinsic parameters from the conditional posterior $p(\boldsymbol{\theta}_{\rm ext}|\boldsymbol{\theta}_{\rm int}, d)$ according to the weights w_i in Eq. (31). The concept and motivation are the same as in Islam *et al.* [51]; sampling Eq. (41) is a lower-dimensional problem than the full posterior $p(\theta|d)$, therefore will typically take less model evaluations to converge robustly. The main difference is that here we include higher modes and precession. Another parameter estimation framework that is based on the marginal likelihood is RIFT [25,46,52,67,68]. RIFT evaluates the marginal likelihood in parallel on a grid over intrinsic parameters, and then constructs a fast interpolator with which it explores the posterior. Our algorithm instead runs on a single core and freshly computes the marginal likelihood at every call. This is rendered possible by the efficiency of our implementation, which computes a marginalized likelihood in $\sim$50 ms. In comparison, RIFT takes tens of seconds on a GPU or minutes on a CPU [52].

We perform two tests of this method. In Sec. IV B 1 we compare it to the more standard strategy of running the sampler on the full parameter space, confirming that we achieve a consistent result on an individual event at a reduced computational cost. In Sec. IV B 2 we generate a large set of synthetic events, and test with P–P plots that the

TABLE I. Configuration of the algorithm used in Sec. IV B.

	Parameter	Value
Δ	Time resolution of the mapping	2^{-13} s
N_j	Number of samples per partial proposal	2048
$N_{\text{eff}}^{\text{min}}$	Minimum effective number of samples	50
j_{max}	Maximum number of proposal adaptations	16
N_ϕ	Number of phase quadrature points	128
	Time series interval around trigger	± 70 ms

method recovers the injected parameters consistently across parameter space.

We perform the parameter estimation runs using the COGWHEEL code [34,69]. We use the IMRPhenomXODE waveform model, which accounts for precession and the $\{(2,2),(2,1),(3,3),(3,2),(4,4)\}$ harmonic modes [70]. The matched-filtering time series $z_{mpd}(t;\boldsymbol{\theta}_{\text{int}})$ and covariances $c_{mm'pp'd}(\boldsymbol{\theta}_{\text{int}})$ are computed using the heterodyne/relative-binning method [71–73], with the implementation of [62]. We "fold" the posterior for the inclination θ_{JN} as described in [34] in order to handle its multimodality (the rest of the parameters identified there as suitable for folding are extrinsic, so the marginalization obviates folding those). We use the stochastic sampler NAUTILUS [64] with 1000 live points. Table I reports the configuration of the marginalization algorithm we used.

1. Comparison to no marginalization

As a first sanity check, we infer the parameters of GW190814 [74]—an event that displays higher modes—in two ways: with our extrinsic marginalization method, or without marginalizing the likelihood and letting the sampler explore the 15-dimensional parameter space. For the non-marginalized case, we fold the $(\theta_{JN}, \hat{\phi}_{\text{net}}, \phi_{\text{ref}}, \psi)$ parameters in order to improve the inference efficiency and robustness [34].

In Fig. 5 we see that the two methods are in excellent agreement on intrinsic and extrinsic parameters, and the likelihood. The run with marginalization took 1.6 h on one CPU core, while the run without marginalization required 5.5 h (3.5 times the cost). Other events gave similar results; using the marginalization was consistently ∼3–5 times faster.

2. Performance on synthetic events

We further assess the performance of the method by means of probability–probability (P–P) plots, shown in Fig. 6. That is, we perform a set of injections on Gaussian noise, with source parameters drawn from a prior distribution. We obtain posterior samples for each injection using the same prior, and test the uniformity of the percentiles P_θ estimated from the posterior samples for various source parameters θ. The percentiles represent the probability that a

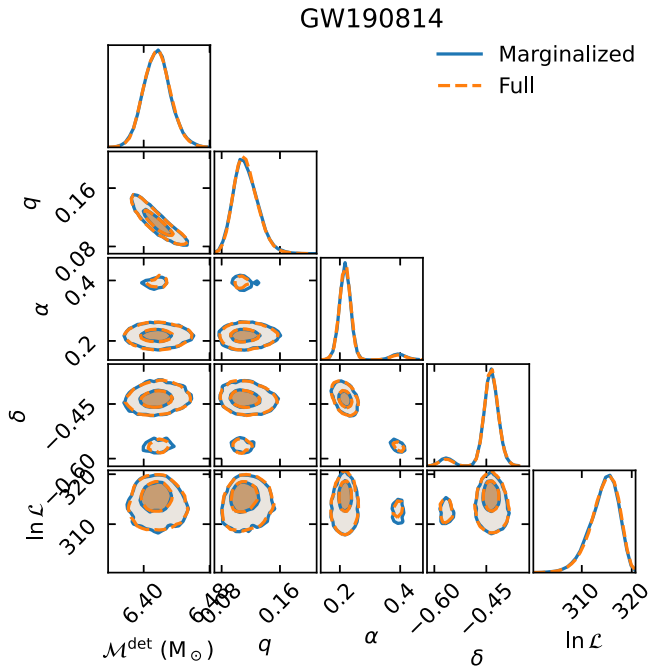

FIG. 5. Application of our method to parameter estimation. In solid blue, we let the stochastic sampler explore the intrinsic parameters following the extrinsic-marginalized posterior, and reconstruct the extrinsic in postprocessing. In dashed orange, we explore all 15 intrinsic and extrinsic parameters with the sampler. Both achieve similar results, but the marginalized case was 3.5 times faster.

parameter lies below the injected value; in a well-calibrated inference they should follow a uniform distribution over the set of injections:

$$P_\theta \equiv \int_{-\infty}^{\theta_{\text{inj}}} \mathrm{d}\theta\, p(\theta|d) \sim \mathrm{U}(0,1). \tag{42}$$

Deviations from uniformity may indicate biases or inaccuracies in our method's performance.

To have more granular information, we partition the parameter space in three bins by detector-frame chirp mass; $\mathcal{M}/M_\odot \in (1,5)$, (5, 25) or (25, 125). Within each bin, we use a mass prior uniform in detector-frame component masses with a cut in mass ratio $q > 1/20$, a "volumetric" spin prior (i.e., isotropic and with $\pi(\chi) \propto \chi^2$ for either dimensionless spin magnitude χ), and uniform in luminosity volume up to $D_{\text{max}} = 1.5$ Gpc in the low-mass bin and 15 Gpc in the other two. To have a sample of events more representative of the set of detections, we further impose a cut $\langle h|h\rangle > 70$ on the injections. We do not use this cut during parameter estimation, but reject samples that do not satisfy it in postprocessing. We use a Hanford-Livingston-Virgo network with average sensitivities from the third observing run. The parameters are specified at a reference frequency of 50 Hz.

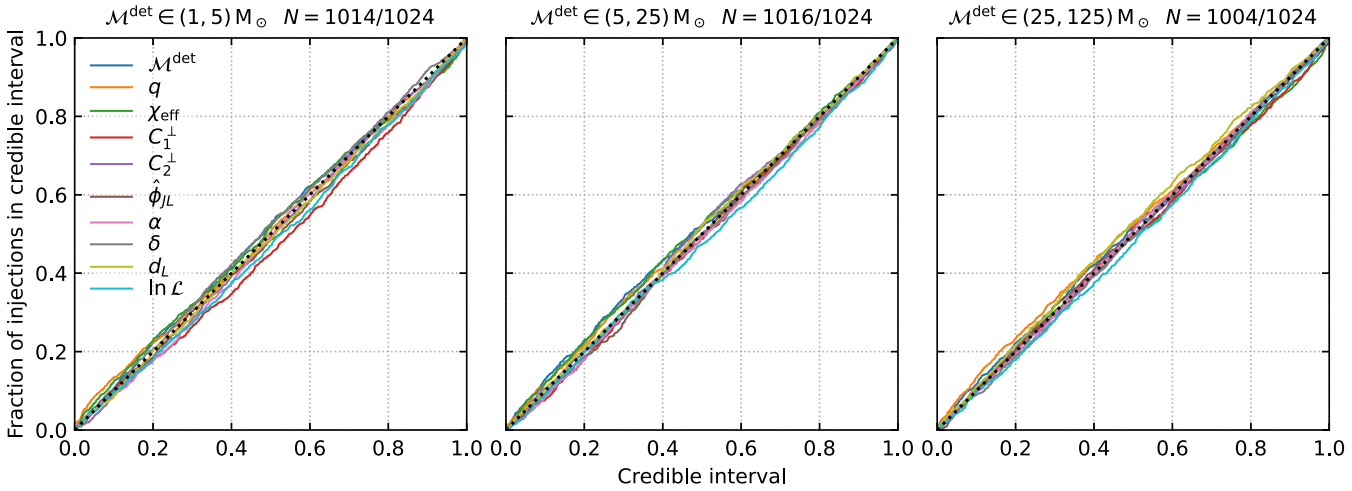

FIG. 6. Probability–probability plots for parameter inference on low-mass, medium-mass and high-mass injections. The empirical distributions of the percentiles are observed to be uniform (their cumulatives follow a diagonal line), indicating satisfactory performance. Each line corresponds to a different source parameter (the parameters are defined in [34]). Titles report the fraction of runs that succeeded on the first attempt; the remainder timed out or failed the $\langle h|h\rangle > 70$ cut, see text. Our method achieves excellent recovery of injection parameters over the wide parameter space that we tested.

We show the results in Fig. 6. 1–2% of the runs failed (timed out at 24 h or produced posteriors below the $\langle h|h\rangle = 70$ cut we imposed on the injection prior), we exclude these from the plot. The fraction of runs that succeeded at the first attempt is reported in the figure titles. We recover satisfactory P–P plots in all parameters and chirp-mass ranges, providing evidence that the posteriors are well-calibrated across the parameter space.

The computational cost of the inferences is summarized in Table II. The average inference runtimes to produce $\sim 10^4$ effective posterior samples range from 0.78 h (high-mass signals) to 2.6 h (low-mass signals) using a single CPU core. Our algorithm is thus reasonably efficient and competitive with other state-of-the-art codes. We find that lower-mass systems take a longer time to run. The reason is threefold; for these systems each waveform generation takes longer, so does the likelihood marginalization, and more likelihood evaluations are needed overall for the sampler to converge. IMRPhenomXODE is slower for lighter systems because this approximant solves a differential equation for the spin dynamics, and these undergo more precession cycles from a given starting frequency if the masses are smaller.[6] One likely explanation for the marginalization being less efficient is that low-mass templates have a shorter autocorrelation length, since these systems emit up to higher frequencies. This allows to measure the arrival time at the detectors better, reducing the target volume of phase space. The reason why the sampler requires more likelihood evaluations might be related to the prominence of various degeneracies in different regions of parameter space.

TABLE II. Timing statistics for the same set of parameter inference on injections shown in Fig. 6. For each chirp-mass range, we report the average inference runtime per event on one CPU core, the average cost $\tau_{\tilde{\mathcal{L}}}$ of each call to the marginalization routine—which dominates that of the waveform model $\tau_{\rm XODE}$—the number $N_{\rm calls}$ of likelihood evaluations performed per event and the average effective sample size achieved (NAUTILUS produces weighted posterior samples).

$\mathcal{M}/M_\odot$	⟨Runtime⟩/h	$\langle\tau_{\tilde{\mathcal{L}}}\rangle$/ms	$\langle\tau_{\rm XODE}\rangle$/ms	$\langle N_{\rm calls}\rangle$	⟨ESS⟩
(1, 5)	2.6	60	23	9.6×10^4	1.2×10^4
(5, 25)	1.2	45	8.2	6.9×10^4	1.2×10^4
(25, 125)	0.78	38	7.1	5.2×10^4	9.7×10^3

The average time for each likelihood marginalization [after the inputs in Eqs. (11) and (13) had been generated] was in the range 38–60 ms (depending on the mass bracket) with $N_{\rm eff}^{\rm min} = 50$; this is in line with the estimation from Fig. 4. The marginalization amounted to approximately 70% of the overall computational cost; unlike the majority of parameter estimation codes, the cost of waveform generation—while not negligible—was not the dominant bottleneck. This suggests that somewhat more expensive models could be used without significantly affecting performance.

In addition to gauging the consistency of our method, the set of injections and posterior samples we generated could

[6]The reason is not simply that the waveforms are longer: this would not occur with analytic approximants such as others in the IMRPhenom family, since the evaluation frequencies are independent of waveform duration in the relative binning algorithm [72].

be utilized for other analyses as a realistic mock catalog of observations [75–78]. With this motivation, we release these data products [79].

C. Low-latency source localization

Beyond marginalizing the likelihood, the weighted samples produced by our algorithm can be used to reconstruct the posterior on extrinsic parameters, including the source location. The computational speed of our marginalization algorithm makes it appealing in the search of short-lived electromagnetic counterparts to gravitational wave signals. Other algorithms in the literature that can localize a source in low-latency are restricted to quadrupolar waveforms [39,43] or high-mass systems that are unlikely to emit light [33]. In contrast, our source localization method works for both high- and low-mass systems and accounts for higher harmonics.

In this section, we will assume that the spins are aligned with the orbit, which allows us to treat the inclination ι as an extrinsic (analytic) parameter and infer its value. We achieve this by including the inclination along with the time, sky location and polarization in the importance sampling (Sec. II D), and including the full spin-weighted spherical harmonics in the waveform model i.e., replacing $e^{im\phi_{\rm ref}}$ by ${}_{-2}Y_{\ell m}(\iota, \phi_{\rm ref})$ in Eq. (7).

In Fig. 7 we demonstrate the application of our algorithm to low-latency source localization. We generate a synthetic signal on Gaussian noise in a Hanford-Livingston-Virgo network with sensitivities typical of the O3 observing run. We simulate the merger between a $1.4M_\odot$ neutron star and a spinning $8M_\odot$ black hole, with dimensionless spin 0.5 aligned with the orbit. Such system could realistically disrupt the neutron star before merger and produce electromagnetic radiation [80]. We place the source in a sky location with good interferometer response at a distance of 100 Mpc, which yields a (recovered) signal-to-noise ratio of 28.6. We simulate the signal using the IMRPhenomXHM approximant [81] (the aligned-spin limit of IMRPhenomXODE).

We retrieve the extrinsic parameter posterior in two ways: modeling the $(\ell, |m|) = \{(2,2), (3,3), (4,4)\}$ harmonics, or only the (2, 2). The latter case is intended to represent the current state of the art in low-latency source localization. At least in this example, we observe that the two results are approximately similar in terms of the sky coordinates. However, higher harmonics provide a significant help for constraining the source inclination, and thereby the distance. (A similar phenomenon was reported in the event GW190412 [53].) This hints at the exciting possibility of ruling out some candidate galaxies in the localization region, facilitating the task of identifying a potential counterpart.

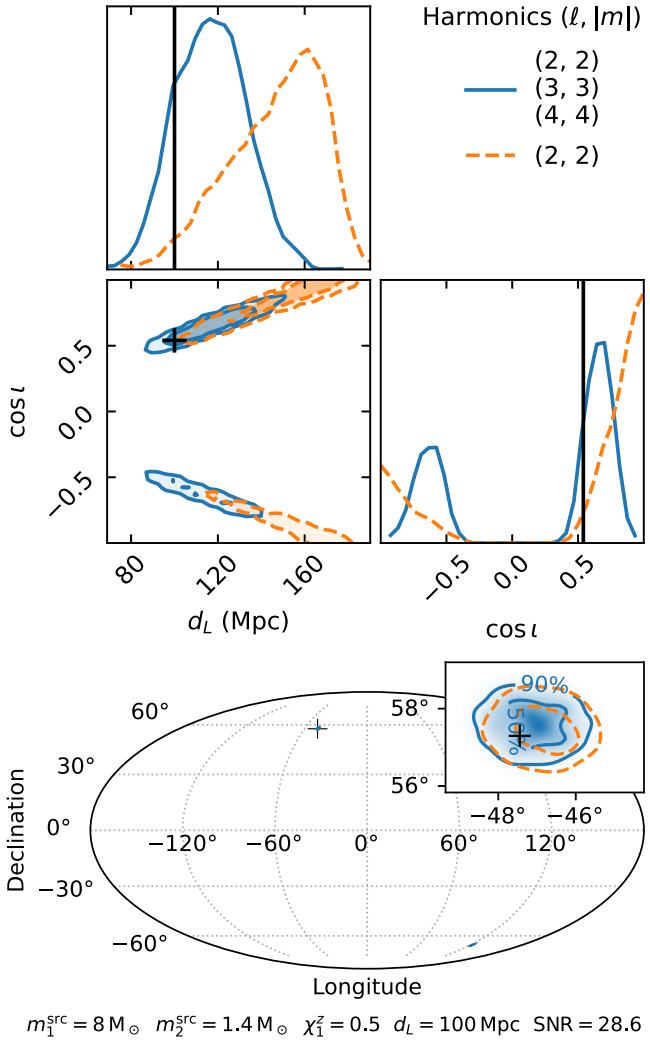

FIG. 7. If intrinsic parameters are available from a search pipeline, our method can localize the source in low latency (which is crucial for multimessenger astronomy) while accounting for higher harmonics. To illustrate this, we show the source location recovery for a synthetic neutron-star–black-hole signal in a Hanford-Livingston-Virgo network in two cases: using a waveform with higher harmonics (solid blue) or without (dashed orange). Injected parameters are indicated with a black cross. In both cases, masses and spins are (unrealistically) set to their true values. Top: higher modes partially lift the distance–inclination degeneracy, improving the low-latency distance measurement relative to current the state of the art, which only includes the quadrupole. Bottom: the source is constrained to two disjoint possible arrival directions, the inset zooms in around the correct one.

Figure 7 was produced with 4000 extrinsic-parameter samples, which took 12 s to generate in a single CPU core (after the matched filtering time series and covariances had been generated those would be provided by a search pipeline). It would be straightforward to produce multiple

smaller batches of samples in parallel if a speedup is desired.

An important caveat in this demonstration is that we treated the intrinsic parameters as known and set them to the true values; in a real application, one would only have access to some noisy estimate from the search pipeline, and would have to marginalize the intrinsic parameters as well. In particular, intrinsic parameters correlate with the distance, as the loudness of the source depends on the masses and spins. This further motivates the inclusion of intrinsic parameters in a more detailed reconstruction of the source location. Higher modes could be of further help in measuring the distance and physical nature of the source, as they also help constrain the intrinsic parameters (by breaking the mass-ratio effective-spin degeneracy). Exploring the consequences of this, as well as the extent to which the inclusion of higher harmonics is important in different regions of parameter space, are left to future work.

V. CONCLUSIONS

We have developed, implemented and tested an efficient algorithm to marginalize the likelihood function of a gravitational wave signal over its extrinsic parameters (and conversely, to sample the posterior conditional on the intrinsic parameters). The computation assumes Gaussian noise and a quasicircular (noneccentric) orbit, and works for signals with precession and/or higher-order harmonics.

For precessing signals (spins misaligned with the orbit), we are able to marginalize out six parameters, namely the orbital phase, distance, coalescence time, polarization angle, right ascension and declination. For aligned-spin signals, we can additionally marginalize the inclination angle. We perform the marginalization over distance via a lookup table, over the orbital phase with trapezoid quadrature, and over the remaining extrinsic parameters using adaptive importance sampling.

Our Python implementation of this algorithm typically achieves a ∼10% accuracy in 50 ms on one CPU core. We make it available through the software COGWHEEL.[7]

We discussed three applications for this tool: search, parameter estimation, and low-latency localization.

In a search for gravitational wave signals, this algorithm is a key piece in the optimal detection statistic, as it computes the Neyman-Pearson likelihood ratio of the hypothesis that there is a signal with given intrinsic parameters versus Gaussian noise. This statistic is computed from the time series of matched-filtered data with the individual harmonic modes of the signal (as opposed to the fully specified waveform), which significantly reduces the size of the template bank [66]. It combines data from different detectors coherently, and correctly penalizes fine-tuned configurations. Our implementation is sufficiently fast that this statistic can be used to rank the large number of foreground and background triggers originating from a search pipeline.

In parameter estimation, by marginalizing the extrinsic parameters we are able to simplify the task of the stochastic sampler; it only needs to explore the intrinsic parameter space, which is lower dimensional and often has a simpler structure. We have demonstrated this on thousands of synthetic signals, recovering satisfactory probability–probability plots across the parameter space. The inference is completed in a one-to-few-hour timescale on a single CPU core, using a waveform model that includes spin-induced precession and higher harmonics.

Finally, we briefly explored the applicability of this algorithm to low-latency source localization, which would be useful in the followup of electromagnetic counterparts to gravitational wave signals. For this application we exploit the capability of efficiently sampling the extrinsic-parameter posterior at given intrinsic parameters. We have shown that accounting for the higher harmonics can make a difference in the recovered distance to the source (by partially lifting the degeneracy with the inclination angle), which suggests the possibility of improving the probability ranking of candidate host galaxies. Interfacing this routine with low-latency search pipelines and demonstrating its performance on synthetic signals are interesting directions for future work.

Beyond these applications, this algorithm could be applied to other use cases with relatively straightforward modifications. For example, the computation of the Bayes factor for a strong gravitational lensing hypothesis given multiple candidate images involves a similar integration of the likelihood over the parameters of the signal [82–85].

ACKNOWLEDGMENTS

We thank Ankur Barsode and Srashti Goyal for helpful discussion. This research has made use of data or software obtained from the Gravitational Wave Open Science Center [86], a service of LIGO Laboratory, the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. LIGO Laboratory and Advanced LIGO are funded by the United States National Science Foundation (NSF) as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through

[7] https://github.com/jroulet/cogwheel.

the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. The construction and operation of KAGRA are funded by Ministry of Education, Culture, Sports, Science and Technology (MEXT), and Japan Society for the Promotion of Science (JSPS), National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea, Academia Sinica (AS) and the Ministry of Science and Technology (MoST) in Taiwan. J. R. acknowledges support from the Sherman Fairchild Foundation. T. V. acknowledges support from NSF Grants No. 2012086 and No. 2309360, the Alfred P. Sloan Foundation through Grant No. FG-2023-20470, the BSF through Award No. 2022136, and the Hellman Family Faculty Fellowship. B. Z. is supported by the ISF, NSF-BSF, a research grant from the Center for New Scientists at the Weizmann Institute of Science and a research grant from the Ruth and Herman Albert Scholarship Program for New Scientists. M. Z. is supported by the Canadian Institute for Advanced Research (CIFAR) program on Gravity and the Extreme Universe and the Simons Foundation Modern Inflationary Cosmology initiative.

APPENDIX A: MISCELLANEOUS COMPUTATIONAL TRICKS

In this technical appendix, we outline steps taken to optimize specific computations, which were omitted from the main text for brevity.

1. Time series relative binning

During the parameter estimation described in Sec. IV B, we generate the matched-filtering time series $z_{mpd}(t;\boldsymbol{\theta}_{\rm int})$ using heterodyne/relative binning, a method that compresses the data to a few hundred weights u by heterodyning with a reference waveform h^0. Thereafter, computing a matched filter requires the waveform evaluated only on a coarse frequency grid $\{f_b\}$. Normally, a single reference waveform in each detector (decomposed by modes) would be used. However, in our application we need to provide $z_{mpd}(t;\boldsymbol{\theta}_{\rm int})$ over a range of ± 70 ms, which means that the test waveform can differ from any single reference by hundreds of autocorrelation times.

Following the formulation of [62], we approximate the matched filtering time series as

$$z_{mpdt} \equiv \int \mathrm{d}f \frac{\tilde{d}_d(f)\tilde{h}^*_{mp}(f)}{S_d(f)} \mathrm{e}^{\mathrm{i}2\pi f t} \tag{A1}$$

$$\approx \sum_{b,m} u_{mtdb} \tilde{h}^*_{mp}(f_b), \tag{A2}$$

$$u_{mtdb} = \frac{1}{\tilde{h}^{0*}_m(f_b)} 4 \int \mathrm{d}f \frac{\tilde{d}_d(f)\tilde{h}^{0*}_{mp}(f)}{S_d(f)} \mathrm{e}^{\mathrm{i}2\pi f t} s_b(f), \tag{A3}$$

where $s_b(f)$ are splines that interpolate the Kronecker delta at the coarse frequency grid,

$$s_b(f_{b'}) = \delta_{bb'}. \tag{A4}$$

The key modification we have made is that we have included the time axis in the summary data, which can be interpreted as using multiple reference waveforms, each with a different time shift. Thus, when generating the time series we are always using an appropriately shifted reference, and we avoid applying a large time shift to the low-resolution waveform. A similar technique had been used in the precursor work of [51].

We note that it is not necessary to use different reference waveforms for different polarizations p, as the ratio $\tilde{h}_{m+}/\tilde{h}_{m\times}$ is typically a smooth function of frequency.

Previous methods have used the fast Fourier transform algorithm to generate the matched filtering time series efficiently [87]. However, that would require waveforms sampled at evenly spaced frequencies, which conflicts with the irregular $\{f_b\}$ used in relative binning.

2. Sparse spline representation

One nuisance associated with having included the time axis in the summary data u_{mtdb} in Eq. (A3) is that now the summary is much larger, to the point that it can require a nontrivial amount of computation.

To mitigate this, we accelerate the frequency integrals in Eq. (A3) (which, in reality, are matrix multiplications along the fine but discrete frequency axis f) by using the B-spline representation of $s_b(f)$. We arrange the splines into a matrix $S_{bf} \coloneqq s_b(f)$, and express it as

$$S_{bf} = \sum_{b'} C_{bb'} B_{b'f}, \tag{A5}$$

where $C_{bb'}$ is a square matrix of coefficients and $B_{b'f}$ is a set of B-splines. The crux is that the B matrix is sparse, which provides a significant speedup. We compute this decomposition using the `SciPy.interpolate.splrep` implementation [88].

Other formulations of the heterodyne/relative binning algorithm [24,72,73,89] work with frequency bins, thus, they do not integrate over the full frequency range and never encounter this problem in the first place. On the other hand, our implementation has the advantage that our approximation of the waveform is smooth over the entire frequency range.

3. Stalling the reference waveform decay

At the core of the heterodyne/relative binning method is the observation that the ratio $\tilde{h}(f)/\tilde{h}^0(f)$ between the test and reference waveforms is a smooth function of frequency. However, a pathological situation can arise when the merger frequency of the reference waveform is lower than that of the test waveform. Then, the ratio diverges at high frequencies and the computation may become numerically unstable. We fix this by using a modified reference waveform in which the high frequency part (where the last 1% of the squared signal-to-noise ratio is accumulated) is set to a nonzero constant, with a smooth cross-fading to prevent artifacts.

4. Waveform time convention

There is a certain amount of arbitrariness in the convention of what is the "arrival time" of a waveform. The practical importance of this for parameter estimation is that the choice of convention can spuriously correlate the time of arrival and the intrinsic parameters (e.g., see Sec. V, [34]). Marginalizing over the arrival time, as done in this and other works, in principle makes this problem moot, since the sampler does not need to deal with those correlations.

However, we did find some cases—especially for highly precessing signals—where the time conventions differed so much that the peak of the matched filtering timeseries $z_{mpd}(t;\boldsymbol{\theta}_{\rm int})$ got shifted by more than our ± 70 ms time window as the intrinsic parameters were varied during parameter estimation. This would cause a complete loss of the signal and bias the inferred parameters. Even in less extreme scenarios, this shift could produce relative-binning errors if the reference and test waveforms are shifted relative to each other.

To fix this, whenever we generate a waveform we apply a time shift to align it to the relative-binning reference. We obtain this time shift from a weighted least-squares linear fit to the unwrapped phase difference $\Delta\Phi$ between the two $m = 2$ waveforms, as follows. We estimate the phase difference as

$$\Delta\Phi(f_b) \coloneqq \mathrm{unwrap}\left[\arg\left(\frac{\tilde{h}_{2+}(f_b)}{\tilde{h}^0_2(f_b)}\right)\right]. \tag{A6}$$

We take the ratio before the argument to ensure that the (potentially very large) phase accumulated by the waveform largely cancels out with that of the reference, rendering the unwrap possible. We apply a weighted least squares linear fit to this phase, with inverse variances

$$\sigma_b^{-2} = \int \mathrm{d}f \, \frac{|\tilde{h}^0(f)| \cdot |\tilde{h}_{2+}(f)|}{S(f)} s_b(f), \tag{A7}$$

where we have defined an effective power spectral density through

$$S^{-1}(f) = \sum_d S_d^{-1}(f). \tag{A8}$$

This procedure maximizes the match between the two waveforms over time and phase, under the approximation that $\tilde{h}_{2+}$ is a small (linear) perturbation of $\tilde{h}^0_2$, and using relative binning to compute the inner product.

We extract the time shift from the slope of the linear fit, and apply it to all the modes and polarizations of the waveform. This ensures that the peak in the time series will occur near that of the reference waveform, and that relative binning errors are kept to a minimum. The constant part of the linear fit plays no role in these two problems, so we discard it. The user is unaffected by this process: in order to facilitate the reconstruction of the signal, we still report the parameter samples in the convention of the original approximant.

5. Memory of previous proposals

During parameter estimation, the likelihood function is evaluated repeatedly at similar parameter values. Hence, it is likely that the adapted proposal $P(\boldsymbol{\tau})$ from one marginalized likelihood call (Sec. II D 4) is also suitable for subsequent calls. With this heuristic, we aim to accelerate the convergence of the importance sampling integral by averaging the initial proposal in each detector $P_d^{(0)}(\tau)$ (as computed in Sec. II D 3) with a "remembered" proposal $P_d^{\rm past}(\tau)$. After each likelihood marginalization call, we update this remembered proposal according to the last iteration of the adaptation,

$$P_d^{\rm past}(\tau) \leftarrow \frac{P_d^{\rm past}(\tau) + \epsilon P_d^{(J)}(\tau)}{1+\epsilon}, \tag{A9}$$

where ϵ is a tunable parameter that controls how fast the remembered proposal is updated; we use $\epsilon = 10^{-2}$.

For this procedure to be effective, it is essential that the time convention preserves the time alignment across waveforms, which we achieve with the method of Sec. A 4.

6. Pruning phases with low maximum likelihood

The phase quadrature in Eq. (21) tries out values of the (very well-measured) orbital phase over its full range. Thus, by construction most of the evaluation points will correspond to waveforms completely inconsistent with the data. We can save many useless computations of $\bar{\mathcal{L}}_D$ by first computing the (cheaper) quantity,

$$2\max_D \ln \mathcal{L} = \frac{\langle d|h\rangle^2}{\langle h|h\rangle}, \tag{A10}$$

and discarding those values of the orbital phase for which the $\max_D \ln \mathcal{L}$ falls short of the maximum one by a large amount (we use $\Delta \max_D \ln \mathcal{L} > 12$).

7. Polarization flip

Before pruning the phases in Appendix A 6, we can salvage the computation invested in some of the points with the following trick. The very worst-fitting orbital phases have a large negative $\langle d|h\rangle$, indicating that the waveform model is in antiphase with the data. In that case, we can improve the proposal at negligible cost by applying a shift of $\pi/2$ to the polarization angle whenever $\langle d|h\rangle < 0$. This operation changes the sign of the antenna coefficients [90], and thereby of the strain h. The transformation reads,

$$\begin{pmatrix} \psi \\ \langle d|h\rangle \\ \langle h|h\rangle \end{pmatrix} \mapsto \begin{pmatrix} \psi + \pi/2 \\ -\langle d|h\rangle \\ \langle h|h\rangle \end{pmatrix}. \tag{A11}$$

This procedure improves N_{eff} to some extent.

8. Foregoing optimization for points with low marginalized likelihood

Especially during the early phase of parameter estimation, the stochastic sampler explores regions of low likelihood and gradually climbs towards the maximum. In those regions, samples are either rejected or heavily downweighted, to the point that they are irrelevant for all practical purposes (i.e., the posterior samples and the Bayesian evidence). While some level of accuracy is desirable, so that the sampler can climb the likelihood surface, for samples with sufficiently low likelihood the target $N_{\text{eff}}^{\min}$ that we impose on the importance sampling integral can be overly conservative.

To save computations in this case, within each inference run we keep track of the maximum recorded value of $\ln \bar{\mathcal{L}}(\boldsymbol{\theta}_{\text{int}})$ up to that point. Whenever the estimated $\ln \bar{\mathcal{L}}$ of the current $\boldsymbol{\theta}_{\text{int}}$ is lower than the historic maximum by a large value (more than 30) we stop optimizing the proposal even if N_{eff} is low.

APPENDIX B: EXAMPLE USAGE

In this appendix we provide a short snippet of code that illustrates how COGWHEEL can be used to estimate the parameters of event GW150914 using the Algorithm described in this article.

```
import matplotlib.pyplot as plt
import pandas as pd
from cogwheel import data
from cogwheel import gw_plotting
from cogwheel import sampling
from cogwheel.posterior import Posterior

# Directory that will contain parameter
# estimation runs:
parentdir = "example"

eventname, mchirp_guess = "GW150914," 30
approximant = "IMRPhenomXPHM"
prior_class = "CartesianIntrinsicIASPrior"

# Download data from GWOSC
filenames, detector_names, tgps = \
      data.download_timeseries(eventname)
event_data = data.EventData.from_timeseries(
      filenames, eventname, detector_names,
  tgps)

# Setup Posterior and Sampler
post = Posterior.from_event(
      event_data, mchirp_guess, approximant,
      prior_class)

sampler = sampling.Nautilus(
      post, run_kwargs = dict (n_live = 1000))

rundir = sampler.get_rundir (parentdir)
sampler.run(rundir) # Will take a while

# Load and plot the samples:
samples = pd.read_feather(
      rundir/sampling.SAMPLES_FILENAME)

gw_plotting.CornerPlot(
      samples, params =
  sampler.sampled_params,
      tail_probability = 1e-4).plot( )
plt.savefig(rundir/f"eventname.pdf,"
      bbox_inches = "tight")
```

[1] B. P. Abbott *et al.*, LIGO: The laser interferometer gravitational-wave observatory, Rep. Prog. Phys. **72**, 076901 (2009).

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

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

[4] S. A. Usman *et al.*, The PyCBC search for gravitational waves from compact binary coalescence, Classical Quantum Gravity **33**, 215004 (2016).

[5] C. Messick *et al.*, Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data, Phys. Rev. D **95**, 042001 (2017).

[6] A. H. Nitz, T. Dent, T. Dal Canton, S. Fairhurst, and D. A. Brown, Detecting binary compact-object mergers with gravitational waves: Understanding and improving the sensitivity of the PyCBC search, Astrophys. J. **849**, 118 (2017).

[7] S. Sachdev *et al.*, The GstLAL search analysis methods for compact binary mergers in Advanced LIGO's second and Advanced Virgo's first observing runs, arXiv:1901.08580.

[8] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, New search pipeline for compact binary mergers: Results for binary black holes in the first observing run of Advanced LIGO, Phys. Rev. D **100**, 023011 (2019).

[9] K. Cannon *et al.*, GstLAL: A software framework for gravitational wave discovery, SoftwareX **14**, 100680 (2021).

[10] T. Dal Canton, A. H. Nitz, B. Gadre, G. S. Cabourn Davies, V. Villa-Ortega, T. Dent, I. Harry, and L. Xiao, Real-time search for compact binary mergers in Advanced LIGO and Virgo's third observing run using PyCBC live, Astrophys. J. **923**, 254 (2021).

[11] B. Zackay, L. Dai, T. Venumadhav, J. Roulet, and M. Zaldarriaga, Detecting gravitational waves with disparate detector responses: Two new binary black hole mergers, Phys. Rev. D **104**, 063030 (2021).

[12] B. Ewing *et al.*, Performance of the low-latency GstLAL inspiral search towards LIGO, Virgo, and KAGRA's fourth observing run, Phys. Rev. D **109**, 042008 (2024).

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

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

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

[16] T. Venumadhav, B. Zackay, J. Roulet, L. Dai, and M. Zaldarriaga, New binary black hole mergers in the second observing run of Advanced LIGO and Advanced Virgo, Phys. Rev. D **101**, 083030 (2020).

[17] S. Olsen, T. Venumadhav, J. Mushkin, J. Roulet, B. Zackay, and M. Zaldarriaga, New binary black hole mergers in the LIGO–Virgo O3a data, Phys. Rev. D **106**, 043009 (2022).

[18] A. K. Mehta, S. Olsen, D. Wadekar, J. Roulet, T. Venumadhav, J. Mushkin, B. Zackay, and M. Zaldarriaga, New binary black hole mergers in the LIGO-Virgo O3b data, arXiv:2311.06061.

[19] D. Wadekar, J. Roulet, T. Venumadhav, A. K. Mehta, B. Zackay, J. Mushkin, S. Olsen, and M. Zaldarriaga, New black hole mergers in the LIGO–Virgo O3 data from a gravitational wave search including higher-order harmonics, arXiv:2312.06631.

[20] A. H. Nitz, C. Capano, A. B. Nielsen, S. Reyes, R. White, D. A. Brown, and B. Krishnan, 1-OGC: The first open gravitational-wave catalog of binary mergers from analysis of public Advanced LIGO data, Astrophys. J. **872**, 195 (2019).

[21] A. H. Nitz, T. Dent, G. S. Davies, S. Kumar, C. D. Capano, I. Harry, S. Mozzon, L. Nuttall, A. Lundgren, and M. Tápai, 2-OGC: Open gravitational-wave catalog of binary mergers from analysis of public Advanced LIGO and Virgo data, Astrophys. J. **891**, 123 (2020).

[22] A. H. Nitz, C. D. Capano, S. Kumar, Y.-F. Wang, S. Kastha, M. Schäfer, R. Dhurkunde, and M. Cabero, 3-OGC: Catalog of gravitational waves from compact-binary mergers, Astrophys. J. **922**, 76 (2021).

[23] J. Veitch *et al.*, Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library, Phys. Rev. D **91**, 042003 (2015).

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

[25] J. Lange, R. O'Shaughnessy, and M. Rizzo, Rapid and accurate parameter inference for coalescing, precessing compact binaries, arXiv:1805.10457.

[26] G. Ashton *et al.*, Bilby: A user-friendly Bayesian inference library for gravitational-wave astronomy, Astrophys. J. Suppl. Ser. **241**, 27 (2019).

[27] G. Ashton and C. Talbot, Bilby-MCMC: An MCMC sampler for gravitational-wave inference, Mon. Not. R. Astron. Soc. **507**, 2037 (2021).

[28] C. M. Biwer, C. D. Capano, S. De, M. Cabero, D. A. Brown, A. H. Nitz, and V. Raymond, PyCBC Inference: A Python-based parameter estimation toolkit for compact binary coalescence signals, Publ. Astron. Soc. Pac. **131**, 024503 (2019).

[29] A. J. K. Chua and M. Vallisneri, Learning Bayesian posteriors with neural networks for gravitational-wave inference, Phys. Rev. Lett. **124**, 041102 (2020).

[30] M. Breschi, R. Gamba, and S. Bernuzzi, Bayesian inference of multimessenger astrophysical data: Methods and applications to gravitational waves, Phys. Rev. D **104**, 042001 (2021).

[31] N. J. Cornish, Rapid and robust parameter inference for binary mergers, Phys. Rev. D **103**, 104057 (2021).

[32] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith, Bayesian parameter estimation using

conditional variational autoencoders for gravitational-wave astronomy, Nat. Phys. **18**, 112 (2021).
[33] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, Real-time gravitational wave science with neural posterior estimation, Phys. Rev. Lett. **127**, 241103 (2021).
[34] J. Roulet, S. Olsen, J. Mushkin, T. Islam, T. Venumadhav, B. Zackay, and M. Zaldarriaga, Removing degeneracy and multimodality in gravitational wave source parameters, Phys. Rev. D **106**, 123015 (2022).
[35] S. Fairhurst, C. Hoy, R. Green, C. Mills, and S. A. Usman, Simple parameter estimation using observable features of gravitational-wave signals, Phys. Rev. D **108**, 082006 (2023).
[36] V. Tiwari, C. Hoy, S. Fairhurst, and D. MacLeod, Fast non-Markovian sampler for estimating gravitational-wave posteriors, Phys. Rev. D **108**, 023001 (2023).
[37] K. W. K. Wong, M. Isi, and T. D. P. Edwards, Fast gravitational-wave parameter estimation without compromises, Astrophys. J. **958**, 129 (2023).
[38] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), Population of merging compact binaries inferred using gravitational waves through GWTC-3, Phys. Rev. X **13**, 011048 (2023).
[39] L. P. Singer and L. R. Price, Rapid Bayesian position reconstruction for gravitational-wave transients, Phys. Rev. D **93**, 024013 (2016).
[40] N. J. Cornish, Rapid and reliable sky localization of gravitational wave sources, arXiv:1606.00953.
[41] S. Sachdev *et al.*, An early-warning system for electromagnetic follow-up of gravitational-wave events, Astrophys. J. Lett. **905**, L25 (2020).
[42] R. Magee *et al.*, First demonstration of early warning gravitational-wave alerts, Astrophys. J. Lett. **910**, L21 (2021).
[43] L. Pathak, S. Munishwar, A. Reza, and A. S. Sengupta, Prompt sky localization of compact binary sources using a meshfree approximation, Phys. Rev. D **109**, 024053 (2024).
[44] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW170817: Observation of gravitational waves from a binary neutron star inspiral, Phys. Rev. Lett. **119**, 161101 (2017).
[45] B. P. Abbott *et al.*, Multi-messenger observations of a binary neutron star merger, Astrophys. J. Lett. **848**, L12 (2017).
[46] C. Pankow, P. Brady, E. Ochsner, and R. O'Shaughnessy, Novel scheme for rapid parallel parameter estimation of gravitational waves from compact binary coalescences, Phys. Rev. D **92**, 023002 (2015).
[47] D. Wadekar, T. Venumadhav, J. Roulet, A. K. Mehta, B. Zackay, J. Mushkin, and M. Zaldarriaga, A new search pipeline for gravitational waves with higher-order modes using mode-by-mode filtering, arXiv: 2405.17400.
[48] S. Dhurandhar, B. Krishnan, and J. L. Willis, Marginalizing the likelihood function for modeled gravitational wave searches, arXiv:1707.08163.
[49] C. Hanna *et al.*, Fast evaluation of multidetector consistency for real-time gravitational wave searches, Phys. Rev. D **101**, 022003 (2020).
[50] L. P. Singer, L. R. Price, B. Farr, A. L. Urban, C. Pankow, S. Vitale, J. Veitch, W. M. Farr, C. Hanna, K. Cannon, T. Downes, P. Graff, C.-J. Haster, I. Mandel, T. Sidery, and A. Vecchio, The first two years of electromagnetic follow-up with Advanced LIGO and Virgo, Astrophys. J. **795**, 105 (2014).
[51] T. Islam, J. Roulet, and T. Venumadhav, Factorized parameter estimation for real-time gravitational wave inference, arXiv:2210.16278.
[52] D. Wysocki, R. O'Shaughnessy, J. Lange, and Y.-L. L. Fang, Accelerating parameter inference with graphics processing units, Phys. Rev. D **99**, 084026 (2019).
[53] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW190412: Observation of a binary-black-hole coalescence with asymmetric masses, Phys. Rev. D **102**, 043015 (2020).
[54] L. S. Finn, Detection, measurement, and gravitational radiation, Phys. Rev. D **46**, 5236 (1992).
[55] LIGO Scientific, Virgo, and KAGRA Collaborations, LVK Algorithm Library—LALSuite, Free software (GPL) (2018).
[56] L. E. Kidder, Using full information when computing modes of post-Newtonian waveforms from inspiralling compact binaries in circular orbit, Phys. Rev. D **77**, 044016 (2008).
[57] G. Pratten, C. García-Quirós, M. Colleoni, A. Ramos-Buades, H. Estellés, M. Mateu-Lucena, R. Jaume, M. Haney, D. Keitel, J. E. Thompson, and S. Husa, Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes, Phys. Rev. D **103**, 104056 (2021).
[58] P. Schmidt, M. Hannam, and S. Husa, Towards models of gravitational waveforms from generic binaries: A simple approximate mapping between precessing and nonprecessing inspiral signals, Phys. Rev. D **86**, 104063 (2012).
[59] K. S. Thorne, Multipole expansions of gravitational radiation, Rev. Mod. Phys. **52**, 299 (1980).
[60] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries, Phys. Rev. D **85**, 122006 (2012).
[61] A. B. Owen, Monte Carlo theory, methods and examples, https://artowen.su.domains/mc/ (2013).
[62] J. Roulet and T. Venumadhav, Inferring binary properties from gravitational-wave signals, Annu. Rev. Nucl. Part. Sci. **74** (2024), 10.1146/annurev-nucl-121423-100725.
[63] B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Chapman & Hall, London, 1986).
[64] J. U. Lange, NAUTILUS: Boosting Bayesian importance nested sampling with deep learning, Mon. Not. R. Astron. Soc. **525**, 3181 (2023).
[65] C. Capano, Y. Pan, and A. Buonanno, Impact of higher harmonics in searching for gravitational waves from nonspinning binary black holes, Phys. Rev. D **89**, 102003 (2014).
[66] D. Wadekar, T. Venumadhav, A. K. Mehta, J. Roulet, S. Olsen, J. Mushkin, B. Zackay, and M. Zaldarriaga, A new approach to template banks of gravitational waves with higher harmonics: Reducing matched-filtering cost by over an order of magnitude, arXiv:2310.15233.
[67] C. A. Rose, V. Valsan, P. R. Brady, S. Walsh, and C. Pankow, Supplementing rapid Bayesian parameter estimation schemes with adaptive grids, arXiv:2201.05263.

[68] J. Wofford, A. B. Yelikar, H. Gallagher, E. Champion, D. Wysocki, V. Delfavero, J. Lange, C. Rose, V. Valsan, S. Morisaki, J. Read, C. Henshaw, and R. O'Shaughnessy, Improving performance for gravitational-wave parameter inference with an efficient and highly-parallelized algorithm, Phys. Rev. D **107**, 024040 (2023).
[69] J. Roulet, S. Olsen, T. Venumadhav, T. Islam, J. Mushkin, and J. Wadekar, JROULET/COGWHEEL: V1.2.1 (2024).
[70] H. Yu, J. Roulet, T. Venumadhav, B. Zackay, and M. Zaldarriaga, Accurate and efficient waveform model for precessing binary black holes, Phys. Rev. D **108**, 064059 (2023).
[71] N. J. Cornish, Fast Fisher matrices and lazy likelihoods, arXiv:1007.4820.
[72] B. Zackay, L. Dai, and T. Venumadhav, Relative binning and fast likelihood evaluation for gravitational wave parameter estimation, arXiv:1806.08792.
[73] N. Leslie, L. Dai, and G. Pratten, Mode-by-mode relative binning: Fast likelihood estimation for gravitational waveforms with spin-orbit precession and multiple harmonics, Phys. Rev. D **104**, 123030 (2021).
[74] R. Abbott *et al.*, Gw190814: Gravitational waves from the coalescence of a 23 solar mass black hole with a 2.6 solar mass compact object, Astrophys. J. Lett. **896**, L44 (2020).
[75] M. Fishbach, D. E. Holz, and W. M. Farr, Does the black hole merger rate evolve with redshift?, Astrophys. J. Lett. **863**, L41 (2018).
[76] T. Callister, M. Fishbach, D. E. Holz, and W. M. Farr, Shouts and murmurs: Combining individual gravitational-wave sources with the stochastic background to measure the history of binary black hole mergers, Astrophys. J. Lett. **896**, L32 (2020).
[77] R. Gray, I. M. Hernandez, H. Qi, A. Sur, P. R. Brady, H.-Y. Chen, W. M. Farr, M. Fishbach, J. R. Gair, A. Ghosh, D. E. Holz, S. Mastrogiovanni, C. Messenger, D. A. Steer, and J. Veitch, Cosmological inference using gravitational wave standard sirens: A mock data analysis, Phys. Rev. D **101**, 122001 (2020).
[78] A. M. Farah, B. Edelman, M. Zevin, M. Fishbach, J. María Ezquiaga, B. Farr, and D. E. Holz, Things that might go bump in the night: Assessing structure in the binary black hole mass spectrum, Astrophys. J. **955**, 107 (2023).
[79] J. Roulet, Auxiliary data release for Fast marginalization algorithm for optimizing gravitational wave detection, parameter estimation and sky localization (Zenodo, 2024), 10.5281/zenodo.10910135.
[80] F. Foucart, T. Hinderer, and S. Nissanke, Remnant baryon mass in neutron star-black hole mergers: Predictions for binary neutron star mimickers and rapidly spinning black holes, Phys. Rev. D **98**, 081501 (2018).
[81] C. García-Quirós, M. Colleoni, S. Husa, H. Estellés, G. Pratten, A. Ramos-Buades, M. Mateu-Lucena, and R. Jaume, Multimode frequency-domain model for the gravitational wave signal from nonprecessing black-hole binaries, Phys. Rev. D **102**, 064002 (2020).
[82] K. Haris, A. K. Mehta, S. Kumar, T. Venumadhav, and P. Ajith, Identifying strongly lensed gravitational wave signals from binary black hole mergers, arXiv:1807.07062.
[83] O. A. Hannuksela, K. Haris, K. K. Y. Ng, S. Kumar, A. K. Mehta, D. Keitel, T. G. F. Li, and P. Ajith, Search for gravitational lensing signatures in LIGO-Virgo binary black hole events, Astrophys. J. Lett. **874**, L2 (2019).
[84] L. Dai, B. Zackay, T. Venumadhav, J. Roulet, and M. Zaldarriaga, Search for lensed gravitational waves including morse phase information: An intriguing candidate in O2, arXiv:2007.12709.
[85] The LIGO Scientific, the Virgo, and the KAGRA Collaborations, Search for gravitational-lensing signatures in the full third observing run of the LIGO-Virgo network, arXiv:2304.08393.
[86] http://www.gwosc.org.
[87] W. Farr, Marginalisation of the time and phase parameters in CBC parameter estimation, https://dcc.ligo.org/public/0114/T1400460/002/margtime.pdf (2014).
[88] P. Virtanen *et al.* (SciPy 1.0 Contributors), SciPy 1.0: Fundamental algorithms for scientific computing in Python, Nat. Methods **17**, 261 (2020).
[89] N. J. Cornish, Heterodyned likelihood for rapid gravitational wave parameter inference, Phys. Rev. D **104**, 104054 (2021).
[90] J. T. Whelan, The geometry of gravitational wave detection, https://dcc.ligo.org/public/0106/T1300666/003/Whelan_geometry.pdf (2013).