Rapid localization and inference on compact binary coalescences with the Advanced LIGO-Virgo-KAGRA gravitational-wave detector network

Soichiro Morisaki,[1,2] Rory Smith,[3,4] Leo Tsukada,[5,6] Surabhi Sachdev,[7,2] Simon Stevenson,[8,4] Colm Talbot,[9,10] and Aaron Zimmerman[11]

[1]*Institute for Cosmic Ray Research, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8582, Japan*
[2]*Leonard E. Parker Center for Gravitation, Cosmology, and Astrophysics, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin 53201, USA*
[3]*School of Physics and Astronomy, Monash University, Victoria 3800, Australia*
[4]*OzGrav: The ARC Centre of Excellence for Gravitational Wave Discovery, Clayton, Victoria 3800, Australia*
[5]*Department of Physics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA*
[6]*Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, Pennsylvania 16802, USA*
[7]*School of Physics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA*
[8]*Centre for Astrophysics and Supercomputing, Swinburne University of Technology, Hawthorn, Victoria 3122, Australia*
[9]*LIGO Laboratory, Massachusetts Institute of Technology, 185 Albany Street, Cambridge, Massachusetts 02139, USA*
[10]*Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA*
[11]*Center for Gravitational Physics, University of Texas at Austin, Austin, Texas 78712, USA*

(Received 16 November 2023; accepted 1 December 2023; published 26 December 2023)

We present a rapid parameter estimation framework for compact binary coalescence (CBC) signals observed by the LIGO-Virgo-KAGRA (LVK) detector network. The goal of our framework is to enable optimal source localization of binary neutron star (BNS) signals in low latency as well as improve the overall scalability of full CBC parameter estimation analyses. Our framework is based on the reduced order quadrature (ROQ) technique and resolves its shortcomings by utilizing multiple ROQ bases in a single parameter estimation run. We have also developed sets of compact ROQ bases for various waveform models, IMRPhenomD, IMRPhenomPv2, IMRPhenomPv2_NRTidalv2, and IMRPhenomXPHM. We benchmark our framework with hundreds of simulated observations of BNS signals by the LIGO-Virgo detector network and demonstrate that it provides accurate and unbiased estimates on BNS source location, with a median analysis time of 6 min. The median searched area is reduced by around 30% compared to estimates produced by BAYESTAR: from 21.8 deg^2 to 16.6 deg^2. Our framework also enables detailed parameter estimation taking into account gravitational-wave higher multipole moments, the tidal deformation of colliding objects, and detector calibration errors of amplitude and phase with the timescale of hours. Our rapid parameter estimation technique has been implemented in one of the LVK parameter estimation engines, BILBY, and is being employed by the automated parameter estimation analysis of the LVK alert system.

DOI: 10.1103/PhysRevD.108.123040

I. INTRODUCTION

The first joint observation of gravitational waves (GWs) and electromagnetic waves from a binary neutron star (BNS) merger has revolutionized relativistic astrophysics [1]. GW emission encoded the dynamics of the colliding objects [2,3], while electromagnetic emission encoded the rich physics of the subsequent short gamma-ray burst [4–6], kilonova [7–10], and afterglow of the merger remnant [11,12]. These complementary observations provided information about the origins of short gamma-ray bursts and heavy elements [13–17], matter with supranuclear densities [18,19], the expansion rate of the Universe [20,21], and the properties of gravity [22–24]. More joint GW and electromagnetic observations—so-called *multimessenger observations*—of compact binary coalescence (CBC) events are required for more accurate and precise understanding of those topics. Increasing the number of the successful multimessenger observations is one of the main goals of the fourth observing

 © 2023 American Physical Society

run (O4) of the LIGO Scientific, Virgo, and KAGRA Collaboration (LVK) [25–28], which is currently ongoing.

Rapid and accurate source localization from GW data is key to successful multimessenger observations of CBC events. Additionally, as the global GW detector network improves in sensitivity and detection rates reach around one per day [29], rapid and accurate parameter estimation on *all* compact binaries ensures that data analysis scales commensurably with increasing detections. Reducing the computational cost of source-parameter estimation has been essential for making rapid and accurate parameter estimation a reality. Several techniques have been developed over the past years to this end. These include likelihood approximation [30–35], parallelized algorithms [36–42], machine learning approaches [43–47], reparametrizations to remove parameter degeneracy and multimodalities [48,49], and other techniques [50–52]. While each of these methods has reduced the wall-time or CPU- or GPU-time cost of parameter estimation (or both) to some degree, they generally require constant updating or modification to reflect progress in, e.g., developments of new model gravitational waveforms; time sensitivity of particular observations (such as observing electromagnetic counterparts); or scalability with increasing event rate. The focus of this work is to present a flexible set of approximate methods for parameter estimation on multiple GW sources for the foreseeable future of LVK observing runs.

Here, we focus on the reduced order quadrature (ROQ) [53,54] method, which accelerates parameter estimation by significantly reducing the amount of waveform evaluations—the dominant run-time cost. The key ingredient in ROQ is a rerepresentation of waveform models as a weighted sum over basis elements and coefficients. The latter contain the waveform's parametric dependence on the CBC's physical parameters, e.g., masses and spins. The smaller the basis size (number of basis elements), the more parameter estimation is accelerated. Previous work [55] has demonstrated that the basis size is drastically reduced if the ROQ basis is constructed over a targeted narrow mass-spin space, reducing the run-time of parameter estimation on BNS to tens of minutes. Parameter estimation provides optimal (in the sense of minimizing the uncertainty) and unbiased sky localization of compact binaries, allowing the odds of discovering an electromagnetic counterpart to be improved if updated sky maps can be quickly disseminated to observers. Analyzing only a restricted region of the mass-spin parameter space may lead to biases in the inference if the data have support outside of the explored region. In addition, the previous work [55] made use of a simple waveform model which does not take into account binary merger dynamics, neutron-star tidal deformability, or generic spin configurations.

In this paper, we present a rapid parameter estimation framework overcoming the shortcomings of the previous approaches, which enables accurate source localization of a BNS signal within minutes and greatly improves the scalability of the detailed parameter estimation analysis, taking into account general binary merger dynamics. The core idea of our framework is to employ multiple ROQ bases in a single parameter estimation analysis: Each basis is constructed in a targeted parameter space to gain a significant speedup, and the union of small patches in parameter space is broad enough to cover the region consistent with observed data. We also present sets of targeted ROQ bases we have developed for use with our optimized framework. Some of the bases have been constructed for computationally cheap waveform models to enable rapid sky localization, and the others for the state-of-the-art waveform models taking into account gravitational-wave higher multipole moments or tidal deformation of colliding objects.

Our rapid parameter estimation technique has been implemented in one of the LVK parameter estimation engines, BILBY [56,57], and that technique as well as our newly developed ROQ bases are being employed by the automated parameter estimation analysis of the LVK O4 alert system, circulating source location estimates to follow-up observers [58–61]. The typical analysis time is several minutes for BNS and hours for the other types of signal. In practice, there can be a delay making the results public due to human vetting of observed data and the inference results. This can increase the time to send out an update General Coordinates Network notice or circular to an hour to several hours. However, in the future, human intervention may be removed from the process so that they are circulated immediately after parameter estimation is completed.

The rest of the paper is organized as follows. In Sec. II, we review the basics of ROQ and describe our optimizations to the ROQ method. In Sec. III, we present our new ROQ bases and describe how they have been constructed. In Sec. IV, we benchmark our optimized ROQ method with hundreds of simulated signals. Finally, in Sec. V, we conclude this paper with summarizing our results. Throughout this paper, we apply the geometric unit system $c = G = 1$.

II. IMPROVED REDUCED ORDER QUADRATURE

In this section, we explain the basics of ROQ and present our idea of using multiple ROQ bases in a single parameter estimation run.

A. Basics

Parameter estimation of a CBC signal is typically based on Bayesian inference, where Bayesian posterior probability density function is computed via Bayes' theorem:

$$p(\theta|\{d_i\}_{i=1}^{N_{\rm det}}) = \frac{\mathcal{L}(\{d_i\}_{i=1}^{N_{\rm det}}|\theta)\pi(\theta)}{\mathcal{Z}}. \tag{1}$$

Here, $\{d_i\}_{i=1}^{N_{\rm det}}$ is a set of data from $N_{\rm det}$ detectors, θ is a set of parameters characterizing the CBC signal, $\mathcal{L}(\{d_i\}_{i=1}^{N_{\rm det}}|\theta)$ is the likelihood function, $\pi(\theta)$ is the prior probability density function, and $\mathcal{Z}$ is evidence. For CBC parameter estimation, we typically assume that instrumental noise is stationary and Gaussian and employ the Whittle likelihood [62], whose logarithm is given by

$$\ln\mathcal{L} = -\frac{1}{2}\sum_{i=1}^{N_{\rm det}}(d_i - h_i(\theta), \quad d_i - h_i(\theta))_i + \text{const} \tag{2}$$

$$= \sum_{i=1}^{N_{\rm det}}\left[(d_i, h_i(\theta))_i - \frac{1}{2}(h_i(\theta), h_i(\theta))_i\right] + \text{const.} \tag{3}$$

$(a,b)_i$ is the noise-weighted inner product,

$$(a,b)_i = \frac{4}{T}\mathrm{Re}\left[\sum_k \frac{a^*(f_k)b(f_k)}{S_i(f_k)}\right], \tag{4}$$

where T is data duration, $S(f)$ is the one-sided power spectral density (PSD) of instrumental noise, and the sum is taken over evenly spaced frequencies $\{f_k\}_k$ ranging from the low-frequency cutoff $f_{\rm low}$ to the high-frequency cutoff $f_{\rm high}$ with the frequency interval of $1/T$. The nonconstant part of $\ln\mathcal{L}$ is referred to as a log-likelihood ratio,

$$\ln\Lambda = \sum_{i=1}^{N_{\rm det}}\left[(d_i, h_i(\theta))_i - \frac{1}{2}(h_i(\theta), h_i(\theta))_i\right], \tag{5}$$

and is computed typically more than millions of times during the stochastic sampling of posterior.

The dominant computational cost of parameter estimation comes from the generation of waveform $\{\tilde{h}_i(f_k;\theta)\}_k$, which is required to compute the log-likelihood ratio (5). The cost is proportional to the number of frequency points $K = (f_{\rm high} - f_{\rm low})T + 1$, which is equal to the number of required waveform evaluations per waveform generation. ROQ reduces the number of required waveform evaluations by expressing the waveform and its squared amplitude as linear functionals of ROQ bases:

$$h_i(f_k;\theta', t_{\rm c} = 0) \simeq \sum_{I=1}^{N_{\rm L}} h_i(F_I;\theta', t_{\rm c} = 0)B_I(f_k), \tag{6}$$

$$|h_i(f_k;\theta)|^2 \simeq \sum_{J=1}^{N_{\rm Q}} |h_i(\mathcal{F}_J;\theta)|^2 C_J(f_k), \tag{7}$$

where $t_{\rm c}$ is the coalescence time of signal, θ' is the set of the parameters except for $t_{\rm c}$, $\{F_I\}_{I=1}^{N_{\rm L}}$ and $\{\mathcal{F}_J\}_{J=1}^{N_{\rm Q}}$ are known as *empirical interpolation nodes*, and $\{B_I(f_k)\}_{I=1}^{N_{\rm L}}$ and $\{C_J(f_k)\}_{J=1}^{N_{\rm Q}}$ as linear and quadratic ROQ bases.[1] Generally, the bases are defined over a subdomain in parameter space. The subdomain is typically smaller than the full parameter space on which the waveform models themselves are defined.

Substituting the above expressions into the log-likelihood ratio (5), one arrives at the compressed ROQ log-likelihood ratio [53,54]:

$$\ln\Lambda_{\rm ROQ} = \sum_{i=1}^{N_{\rm det}}\left[L_i(\theta) - \frac{1}{2}Q_i(\theta)\right], \tag{8}$$

where the functions $L_i(\theta)$ and $Q_i(\theta)$ are given, respectively, by

$$L_i(\theta) = \mathrm{Re}\left[\sum_{I=1}^{N_L} h_i(F_I;\theta', t_{\rm c} = 0)\omega_{I,i}(t_c)\right], \tag{9}$$

$$Q_i(\theta) = \sum_{J=1}^{N_Q} |h_i(\mathcal{F}_J;\theta)|^2\psi_{J,i}. \tag{10}$$

The quantities $\omega_{I,i}(t_c)$ and $\psi_{J,i}$ are integration weights that depend only on the bases, data, and noise power spectral density:

$$\omega_{I,i}(t_c) = \frac{4}{T}\sum_k \frac{d_i^*(f_k)B_I(f_k)}{S_i(f_k)}e^{-2\pi i f_k t_c}, \tag{11}$$

$$\psi_{J,i} = \frac{4}{T}\sum_{k=1}^{K}\frac{C_J(f_k)}{S_i(f_k)}. \tag{12}$$

These data-dependent weights are a one-time, upfront calculation and can be efficiently computed using an inverse fast Fourier transform. Since $\ln\Lambda_{\rm ROQ}$ can be computed with waveform values at $N_{\rm L} + N_{\rm Q}$ frequency points, the number of required waveform evaluations is reduced by $K/(N_{\rm L} + N_{\rm Q})$, and the analysis is expected to be accelerated by the same factor.

The ROQ bases $\{B_I(f_k)\}_{I=1}^{N_{\rm L}}$ and $\{C_J(f_k)\}_{J=1}^{N_{\rm Q}}$ need to be preconstructed and stored. For signal with long duration, which has large K, their file sizes can be a few tens of gigabytes or even larger. This gets a more serious issue when tens or hundreds of bases are constructed for different mass-spin subdomains, as we do in this work. This practical issue can be resolved by utilizing the likelihood approximation technique developed in [34]. In this approximation, the total frequency range is divided into B frequency bands with a set of smooth window functions $\{w^{(b)}(f)\}_{b=1}^{B}$. They are constructed so that signal from the starting frequency of the bth band has duration shorter than a

[1]Note that the representations in Eq. (6) are often referred to as "reduced order models" or ROMs of waveforms. Here, we choose to avoid using the term ROM to minimize the amount of technical jargon, as we are primarily interested in the quantities derived from Eq. (6).

certain duration value $T^{(b)}$, and their values can be chosen so that they are decreasing $T = T^{(1)} > T^{(2)} > \cdots > T^{(B)}$ thanks to the chirping nature of CBC signal (the increase of frequency with time). The start and end frequencies of the frequency bands are determined based on the time-frequency relation of CBC signal computed with a certain value of detector-frame chirp mass.

Then, the frequency sum of $(d_i, h_i(\theta))_i$ is decomposed into sums over the B frequency bands, and the sum over the bth band is approximately computed with $h_i(f;\theta)$ at downsampled frequency points $f_k^{(b)} = k/T^{(b)}$. Similarly, $(h_i(\theta), h_i(\theta))_i$ is decomposed, and the sum over the bth band is approximately computed with $|h_i(f;\theta)|^2$ at $\hat{f}_k^{(b)} = k/\hat{T}^{(b)}$, $\hat{T}^{(b)} = \min[2T^{(b)}, T]$. Since $1/T^{(b)} \gg 1/T$ for large b, this *multiband* approximation significantly reduces the number of required waveform values at high frequency. Substituting Eqs. (6) and (7) into the multiband forms of the inner products, they are approximated by ROQ bases at the downsampled frequency points [$B_I(f)$ at $\{\{f^{(b)}\}_k\}_{b=1}^B$ and $C_J(f)$ at $\{\{\hat{f}^{(b)}\}_k\}_{b=1}^B$]. The exact forms of the multibanded ROQ inner products are given in Appendix A. Hence, we need to store only the multibanded basis components. For typical BNS signal in the LVK frequency range, the original number of frequency points $K \sim 10^6$ is reduced to $\sim 10^4$ with the multiband approximation [34], and, hence, the file size is reduced by a factor of $\sim$100.

B. Using multiple ROQ bases for arbitrary mass-spin priors

In previous parameter estimation analyses using ROQ likelihoods, the explored parameter range has typically been set by the width of a single ROQ mass-spin partition. Hence, the width of ROQ bases has been designed to be wide enough so that posterior distributions of a signals' mass and spin will be comfortably contained within them. This introduces a trade-off between efficiency and accuracy: ROQ basis constructed over a narrow mass-spin space has a smaller basis size but restricts the explored parameter space.

Besides this trade-off, there are a number of drawbacks to this approach. Analysts often chose priors based on astrophysical considerations: These might be wider than those offered by individual ROQ bases; a detection trigger might have masses which fall near the boundary of an ROQ basis; diagnostic checks (such as p-p tests) may require a consistent broad prior; and catalogs of gravitational-wave events might want to impose consistent priors for particular classes of events.

To overcome these issues, it is straightforward to define an ROQ likelihood over a parameter domain larger than the small subdomains of the individual ROQ basis. One simply builds multiple *sets* of ROQ weights constructed from bases covering different parameter subdomains. The likelihood over the full domain is then just the union of the likelihoods on the individual subdomains:

$$\mathcal{L}_{\rm ROQ}(d|\theta, \text{all bases}) = \begin{cases} \mathcal{L}_{\rm ROQ}(d|\theta, \text{particular basis}), & \text{for } \theta \text{ in domain of a particular basis,} \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

Thus, when a parameter sample θ is drawn, one simply computes the likelihood function using the precomputed ROQ weights associated with the basis set whose domain contains θ. This likelihood construction from multiple ROQ bases has been implemented in ROQGravitationalWaveTransient of BILBY. In the current implementation, ROQ bases used in a single run are assumed to have the same values of $f_{\rm low}$, $f_{\rm high}$, and $1/T$, while it can, in principle, be generalized.

III. CONSTRUCTION OF REDUCED ORDER QUADRATURE BASES

Based on the idea of using multiple ROQ bases, we have constructed sets of ROQ bases constructed over targeted mass-spin subdomains. In this section, we present those bases as well as explaining how they have been constructed. In this work, we consider dividing the parameter space based on a single parameter, detector-frame chirp mass[2] $\mathcal{M} \equiv (m_1 m_2)^{3/5}/(m_1 + m_2)^{1/5}$, where m_1 and m_2 are detector-frame component masses satisfying $m_1 \geq m_2$. This parameter is known to predominantly determine the gravitational waveforms' phase and amplitude evolution and groups waveforms with similar morphologies. Using more sophisticated parameters such as the linear combinations of GW phase coefficients introduced by [55] may reduce the basis sizes further, but we leave exploration in that direction to future work.

A. Waveforms and parameter ranges

We consider three different mass ranges: BNS, binary black hole (BBH), and intermediate regions. For each region, we have constructed ROQ bases for different waveform models and parameter ranges. All of our bases have been constructed to approximate waveform values from $f_{\rm low} = 20$ Hz, the default low-frequency cutoff used in LVK parameter estimation analyses [63,64]. They are summarized in Table I.

1. BNS

Astrophysical BNS masses plausibly span a range between around 1 and $2M_\odot$, with mass ratios $q \equiv m_2/m_1$

[2]Sometimes referred to as redshifted chirp mass.

TABLE I. Waveform models and parameter ranges for which ROQ bases have been constructed and widths of chirp-mass subdomains.

Mass range	Waveform model	Spin range	Tides range	Subdomain width
BNS ($0.6M_\odot \leq \mathcal{M} \leq 4.0M_\odot, 1/8 \leq q \leq 1$)	IMRPhenomD	$a \leq 0.05$	⋯	$\Delta\mathcal{M}^{-5/3} = 0.01$ ($M_\odot^{-5/3}$)
	IMRPhenomPv2	$a \leq 0.99$	⋯	$\Delta\mathcal{M}^{-5/3} = 0.1$ ($M_\odot^{-5/3}$)
	IMRPhenomPv2_NRTidalv2		$\Lambda \leq 5000$	
Intermediate ($1.4M_\odot \leq \mathcal{M} \leq 21M_\odot, 1/20 \leq q \leq 1$)	IMRPhenomPv2	$a \leq 0.99$	⋯	$\Delta\mathcal{M}^{-5/3} = 0.01$ ($M_\odot^{-5/3}$)
BBH ($10.02M_\odot \leq \mathcal{M} \leq 200M_\odot, 1/20 \leq q \leq 1$)	IMRPhenomXPHM	$a \leq 0.99$	⋯	See Sec. III C

roughly in the range $0.5 \leq q \leq 1$ [65]. The upper and lower limits are uncertain due to a limited number of galactic and extragalactic observations, together with incomplete models of binary neutron star astrophysics. The dimensionless spin magnitudes of colliding neutron stars, a_1 and a_2, can be up to ~0.7 assuming a plausible equation of state of matter with supranuclear densities [66], while BNS systems that have been found by electromagnetic observations and will merge within a Hubble time will have spins of 0.04 at largest when they merge [67]. Tidal deformation of the stars also affects the gravitational waveform [68], and this effect is characterized by dimensionless tidal deformability parameters Λ_1 and Λ_2. Measurements of those values provide constraints on the uncertain equation of state of matter with supranuclear densities [18,19] and are of particular interest for nuclear physics.

Our BNS ROQ bases span detector-frame chirp masses in the range $0.6M_\odot \leq \mathcal{M} \leq 4M_\odot$ and mass ratios in the range $1/8 \leq q \leq 1$. While this mass range extends far beyond what is plausible for astrophysical BNSs, our motivation is to provide an "insurance buffer" to accommodate unexpected sources, unusually broad posterior densities, etc. Over the mass range, we have constructed ROQ bases for three different sets of waveform models and parameter ranges: IMRPhenomD [69,70] for the low-spin range $0 \leq a_1, a_2 \leq 0.05$, IMRPhenomPv2 [71] for the high-spin range $0 \leq a_1, a_2 \leq 0.99$, and IMRPhenomPv2_NRTidalv2 [72] for the high-spin range and the broad tidal deformability range $0 \leq \Lambda_1, \Lambda_2 \leq 5000$. IMRPhenomPv2 and IMRPhenomPv2_NRTidalv2 have cusps in waveform at a certain mass-spin space, where $q \lesssim 0.4$ and spins are antialigned with the orbital angular momentum. As discussed in [54], those cusps make it practically impossible to obtain converged ROQ bases. Since the waveform models are not valid in that region anyway, we exclude that parameter space for basis construction. The excluded mass ratio-spin region for $\mathcal{M} = 1M_\odot$ is shown in gray in Fig. 1, where χ_1 and χ_2 are spin components projected onto the orbital angular momentum, and the region is almost unchanged for a different $\mathcal{M}$ value within the range. More details about the waveform cusps will be explained in Appendix B.

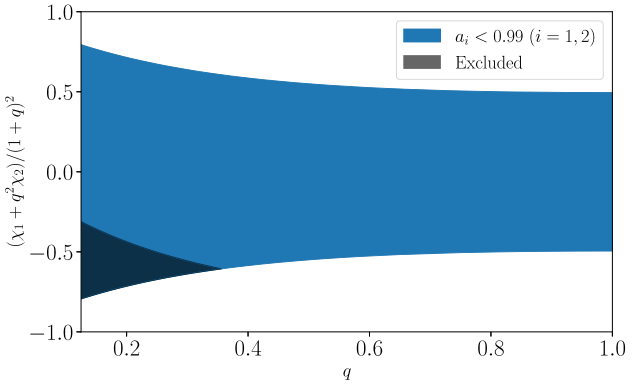

FIG. 1. Excluded region of mass ratio q and projected total spin $\chi = (\chi_1 + q^2\chi_2)/(1+q)^2$ in the basis construction of IMRPhenomPv2 and IMRPhenomPv2_NRTidalv2, where χ_1 and χ_2 are dimensionless spins projected onto the orbital angular momentum. The gray region is the excluded region, and the blue region is the region allowed by the spin limit $a_1, a_2 \leq 0.99$. The excluded region is determined by Eq. (B4), and the gray region presented here is calculated with $\mathcal{M} = 1M_\odot$ and $f_{\text{high}} = 4096$ Hz, while it very weakly depends on $\mathcal{M}$ within the range we consider, $0.6M_\odot \leq \mathcal{M} \leq 4.0M_\odot$.

The sole purpose of the low-spin IMRPhenomD ROQ bases is to provide rapid sky location for BNS candidates. IMRPhenomD is valid only for simple spin configurations where spins are aligned with the orbital angular momentum. While this waveform restriction and the low-spin assumption does not allow us to explore a broader, more agnostic spin space, parameter estimation using those bases is extremely quick, enabling us to provide estimated sky location with the timescale of minutes as demonstrated in Sec. IV. On the other hand, IMRPhenomPv2 is valid for general spin configurations and IMRPhenomPv2_NRTidalv2 also takes into account tidal deformation of colliding objects. Those two bases are useful for more detailed follow-up analysis of BNS candidates.

2. BBH

For the BBH mass range, we have constructed ROQ bases for IMRPhenomXPHM [73], which is valid for general spin configurations and takes into account GW higher multipole moments. Astrophysical BBH masses observable by LVK detectors span a range between around 2 and $400M_\odot$. The mass ratio distribution is subject to large uncertainties; however, the IMRPhenomXPHM

model itself accurately describes binaries with mass ratios in the range $1/20 \lesssim q \leq 1$. Current constraints on the spin magnitude and orientations of astrophysical BBHs allow any magnitude up to 1 and any possible orientation [74]. We allow our bases to span the spin magnitude range $0 \leq a_1, a_2 \leq 0.99$ and all possible spin directions.

Currently, the mass range of our bases spans detector-frame chirp masses in the range $10.02M_\odot \leq \mathcal{M} \leq 200M_\odot$ and mass ratios in the range $1/20 \leq q \leq 1$. The mass range does not extend to the lowest mass regions of the binary black hole space, which is due to current technical limitations in the design and construction of ROQ bases. Specifically, building comprehensive training sets for long-duration, low-mass binary black hole waveforms is more challenging than for binary neutron stars because of the presence of higher multipole moments in the signals. This requires significantly larger training sets, and hence memory, than is currently feasible. Constructing low-mass IMRPhenomXPHM ROQ bases will be the subject of future work.

3. Intermediate

To fill in the gap between the BNS and BBH mass ranges, we have constructed ROQ bases of IMRPhenomPv2 for detector-frame chirp masses in the range $1.4M_\odot \leq \mathcal{M} \leq 21M_\odot$. Our bases span mass ratios in the range $1/20 \leq q \leq 1$, which is broad enough to include any neutron star–black hole (NSBH) binaries that may lead to electromagnetic counterparts [75], and spins in the range $0 \leq a_1, a_2 \leq 0.99$. As with the BNS bases, we exclude the mass-spin space where the waveform has cusps. These bases are being used for automated LVK parameter estimation of event candidates which do not fall in the BNS or BBH mass region.

B. Mass-frequency partitions and multibanding

Following [54], we divide each mass region into several overlapping chirp-mass partitions with different frequency resolution $1/T$ and high-frequency cutoffs $f_{\rm high}$. The chirp-mass range and T of each partition are determined so that the ROQ bases in that partition can accurately model any waveform whose duration falls between 2^{n-1} and 2^n s, where n is an integer. More mathematically, given a map from waveform duration τ to $\mathcal{M}$, $\mathcal{M}(\tau)$, the chirp-mass range is determined by

$$\mathcal{M}((2^n - 2.1)\ \mathrm{s}) \leq \mathcal{M} \leq 1.2\mathcal{M}((2^{n-1} - 2.1)\ \mathrm{s}), \quad (14)$$

and $T = 2^n$ s. The time offset of 2.1 s is to accommodate the time between the coalescence time $t_{\rm c}$ and the end time of analyzed data $t_{\rm e}$. Let $t_{\rm trig}$ be trigger time reported at signal detection, we typically have 2 s safety margin between $t_{\rm trig}$ and $t_{\rm e}$, and the standard prior of $t_{\rm c}$ is uniform distribution in $t_{\rm trig} - 0.1\ \mathrm{s} \leq t_{\rm c} \leq t_{\rm trig} + 0.1$ s [76]. Hence, $(2 + 0.1)$ s is the maximum time difference between $t_{\rm c}$ and $t_{\rm e}$ in the standard parameter estimation, ignoring the light-traveling time between the geocenter and a detector.

For the BNS and intermediate mass regions, we employ the leading-order $\tau - \mathcal{M}$ relation:

$$\mathcal{M}(\tau) = \left(\frac{5}{256\pi f_{\rm low}\tau}\right)^{\frac{3}{5}} \frac{1}{\pi f_{\rm low}}, \quad (15)$$

with $f_{\rm low} = 20$ Hz. With this algorithm, the BNS mass region is divided into four chirp-mass partitions: $0.6M_\odot \leq \mathcal{M} \leq 1.1M_\odot$, $0.92M_\odot \leq \mathcal{M} \leq 1.7M_\odot$, $1.4M_\odot \leq \mathcal{M} \leq 2.6M_\odot$, and $2.1M_\odot \leq \mathcal{M} \leq 4.0M_\odot$, with $T = 512$, 256, 128, and 64 s, respectively. For the low-spin IMRPhenomD bases, $f_{\rm high}$ is 1024 Hz for all the partitions, which is high enough not to degrade estimates on source locations. For the IMRPhenomPv2 and IMRPhenomPv2_NRTidalv2 bases, $f_{\rm high}$ is 4096 Hz for the first three partitions and 2048 Hz for the last partition to incorporate all the information on binary merger. The intermediate mass region is divided into five partitions: $1.4M_\odot \leq \mathcal{M} \leq 2.6M_\odot$, $2.1M_\odot \leq \mathcal{M} \leq 4.0M_\odot$, $3.3M_\odot \leq \mathcal{M} \leq 6.3M_\odot$, $5.2M_\odot \leq \mathcal{M} \leq 11M_\odot$, and $8.7M_\odot \leq \mathcal{M} \leq 21M_\odot$ with $T = 128$, 64, 32, 16, and 8 s, respectively. $f_{\rm high}$ is 1024 Hz for all the partitions.

IMRPhenomXPHM contains GW higher multipole moments and their frequency-time relation is different from that of the dominant quadrupole moment. However, the duration of the $(l, |m|)$ modes can approximately be calculated by the same formula (15) with the frequency scaling $f_{\rm low} \to (2/|m|)f_{\rm low}$. To obtain the most conservative value of T, we assume $|m| = 4$, the highest $|m|$ of IMRPhenomXPHM leading to the longest duration, and employ the relation (15) with $f_{\rm low} = 10$ Hz. With this relation, the BBH region is divided into four partitions: $10.02M_\odot \leq \mathcal{M} \leq 19.05M_\odot$, $17.32M_\odot \leq \mathcal{M} \leq 31.85M_\odot$, $26.54M_\odot \leq \mathcal{M} \leq 62.86M_\odot$, and $52.38M_\odot \leq \mathcal{M} \leq 200.0M_\odot$ with $T = 32$, 16, 8, and 4 s, respectively.

To reduce the total file size of ROQ bases with large T, we downsample all the BNS bases, and the NSBH bases with $T = 128$, 64, and 32 s using the multiband approximation. The frequency bands of each partition are determined based on the time-frequency relation calculated with the chirp-mass value $0.95\mathcal{M}_{\rm min}$, where $\mathcal{M}_{\rm min}$ is the minimum chirp mass of the partition and 0.95 is a safety factor. The duration of the band decreases at the rate of $1/2$, $T^{(b)} = T/2^{b-1}$.

C. Chirp-mass subdomains

Each partition is further divided into narrow $\mathcal{M}$ subdomains to reduce the basis sizes. Each partition of the BNS and intermediate mass regions is divided equally in $\mathcal{M}^{-5/3}$, the leading-order mass combination entering into GW phasing. For the IMRPhenomPv2 and IMRPhenomPv2_NRTidalv2 bases of the BNS mass region, each partition is

divided into subdomains with the width of $\Delta(\mathcal{M}^{-5/3}) = 0.1M_{\odot}^{-5/3}$, and one set of linear and quadratic bases have been constructed over each subdomain. It results in 15, eight, four, and two linear and quadratic bases for the partitions of $T = 512$, 256, 128, and 64 s, respectively. For the low-spin IMRPhenomD bases, each partition is divided into subdomains with the width of $\Delta(\mathcal{M}^{-5/3}) = 0.01M_{\odot}^{-5/3}$, narrower subdomains to obtain compact bases for low-latency source localization. We have confirmed that reducing the chirp-mass width further reduces the basis sizes only by a few tens of percent. Only the linear basis is constructed over each subdomain. Conversely, since the quadratic basis of low-spin IMRPhenomD does not significantly depend on the width of a subdomain and its basis size is much smaller than the sizes of the linear bases, the quadratic bases have been constructed over the whole mass partitions. It results in 149, 74, 37, and 20 linear bases for the partitions of $T = 512, 256$, 128, and 64 s, respectively, and one quadratic basis per partition.

For the IMRPhenomPv2 bases of the intermediate mass region, each partition is divided into subdomains with the width of $\Delta(\mathcal{M}^{-5/3}) = 0.01M_{\odot}^{-5/3}$, and one set of linear and quadratic bases have been constructed over each subdomain. It results in 37, 20, ten, five, and three linear and quadratic bases for the partitions of $T = 128$, 64, 32, 16, and 8 s, respectively.

For IMRPhenomXPHM, we take a somewhat more *ad hoc* approach than for the other waveform models used in this paper. Studies to optimize the widths of subdomains are still ongoing, and we present the results for the bases which are currently being used in the LVK automated parameter estimation analysis. The 4, 8, and 16 s mass spaces are split into 11 equally sized subdomains, chosen to manage memory and computational resources. The 32 s mass space is split into 24 equal subdomains. We have found empirically that it yields comparably sized base sets, though we note that they are not optimal in the sense that further reduction in size will likely yield more compact sets. This is the topic of future work.

D. Basis construction

For constructing IMRPhenomXPHM bases, we employ the same strategy and code base as used in [54]. The basis construction is a combination of the *greedy algorithm* [77] and the *empirical interpolation method*. First, the greedy algorithm is run on a set of randomly drawn CBC waveforms, and N reduced basis vectors are obtained, whose span can approximate any waveform in the set within a specified accuracy. The set of waveforms used for constructing the reduced basis is referred to as *training set*. Next, the empirical interpolation method uses the N reduced basis vectors to construct N empirical interpolation nodes and ROQ basis. The interpolant is then *validated* by computing its representation errors for waveforms outside the training set. If there are waveforms whose errors exceed an error tolerance, these waveforms are added to the training set, and the whole process is repeated. In the IMRPhenomXPHM model, waveform morphology is determined by the ten parameters, two masses, two spin vectors, the inclination angle of the orbital plane, and the coalescence phase. Hence, the training and validation sets consist of waveforms with random realizations of those ten parameter values.

For the BNS and NSBH bases, we employ a slightly different strategy. We skip the empirical interpolation method in the iterative loop and validate reduced basis vectors by computing their projection errors for waveforms outside the training set. Once reduced basis vectors passing the validation test are obtained, they are mapped to empirical nodes and ROQ basis with the empirical interpolation method.

Amplitude and phase deviations due to detector calibration errors [78] are also taken into account for the BNS and NSBH bases. Those deviation factors are randomly realized and multiplied by a certain fraction of waveforms in the training and validation sets. They are calculated via spline interpolation of their values at ten nodal frequency points distributed log-uniformly. Their values at the nodes are drawn from uniform distribution from -20% to 20% for amplitude and $-15°$ to 15° for phase. While they are not considered in the basis construction for IMRPhenomXPHM, the bases are shown to be accurate to represent waveforms incorporating those deviations in Sec. III F.

For IMRPhenomD, GW polarizations $h_+(f)$ and $h_\times(f)$ are linearly dependent. Hence, the reduced bases need to be constructed only for $h_+(f)$ and $|h_+(f)|^2$. Its waveform morphology is parametrized by the four parameters $(m_1, m_2, \chi_1, \chi_2)$. The initial training set consists of $3 \times 5^4 = 1875$ waveforms, where one-third of them are waveforms on 5^4 grid points of $\mathcal{M} - q - \chi_1 - \chi_2$ space with vanishing calibration errors and two-thirds of them are waveforms on the same mass-spin grid points with random calibration errors. The reduced basis is validated against 3×10^4 waveforms, where one-third of them are with vanishing calibration errors and two-thirds of them are with random calibration errors.

As explained in a previous subsection, all the IMRPhenomD bases are multibanded. Rather than constructing ROQ basis for fully sampled waveforms and downsampling it, we downsample the waveforms in the training and validation sets and directly construct multibanded ROQ basis. To reflect the integration weights in the multiband inner products, (A1) and (A2), we also multiply $\sqrt{w^{(b)}(f_k^{(b)})/T^{(b)}}$ and $\sqrt{w^{(b)}(\hat{f}_k^{(b)})/\hat{T}^{(b)}}$ by each component of waveforms for linear and quadratic basis construction, respectively. The waveforms are normalized in the following ways for linear and quadratic basis construction, respectively:

$$\sum_{b=1}^{B}\frac{1}{T^{(b)}}\sum_{k}w^{(b)}(f_k^{(b)})|h_+(f_k^{(b)})|^2=1, \tag{16}$$

$$\sum_{b=1}^{B}\frac{1}{\hat{T}^{(b)}}\sum_{k}w^{(b)}(\hat{f}_k^{(b)})|h_+(\hat{f}_k^{(b)})|^2=1. \tag{17}$$

The error tolerances for projection errors are 10^{-10} for all the linear bases and 6.4×10^{-15}, 9.8×10^{-15}, 1.4×10^{-14}, and 1.7×10^{-14} for quadratic bases with $T=512$, 256, 128, and 64 s, respectively. Those tolerance values are empirically determined so that the relative log-likelihood-ratio errors measured in the next subsection are $\lesssim10^{-4}$.

For IMRPhenomPv2 or IMRPhenomPv2_NRTidalv2, $h_+(f)$ and $h_\times(f)$ are linear combinations of the following five base waveforms (See Appendix C for more details):

$$l_m(f)=\mathrm{e}^{\mathrm{i}(m\alpha(f)-2\epsilon(f))}d^2_{2,m}(-\beta(f))h_\mathrm{D}(f)\quad(m=-2,-1,0,1,2), \tag{18}$$

where $\alpha(f)$, $\beta(f)$, and $\epsilon(f)$ are Euler angles to parametrize the rotation from an inertial frame whose z axis is aligned with the total angular momentum to a coprecessing frame whose z axis is aligned with the orbital angular momentum, $d^l_{m',m}(\beta)$ is the component of the Wigner matrix, and $h_\mathrm{D}(f)$ is the waveform in the coprecessing frame computed with the IMRPhenomD model. The basis vectors are constructed to approximate those five base waveforms instead of the original polarizations. The base waveforms are parametrized by the five parameters $(m_1,m_2,\chi_1,\chi_2,\chi_\mathrm{p})$ for IMRPhenomPv2 and the seven parameters $(m_1,m_2,\chi_1,\chi_2,\chi_\mathrm{p},\Lambda_1,\Lambda_2)$ for IMRPhenomPv2_NRTidalv2, where χ_p is the effective precessing spin parameter [79]. They do not include the angle between the line of sight and the total angular momentum, θ_J, and the initial phase of $\alpha(f)$, α_0, while those two parameters need to be taken into account when basis vectors are constructed for the original polarizations [54]. Hence, building the basis vectors for the base waveforms reduces the number of parameters by 2.

On the other hand, $|F_+h_+ + F_\times h_\times|^2$, where F_+ and $F_\times$ are detector beam pattern functions, is the linear combination of the following base waveforms:

$$q^\mathrm{cos}_{m,m'}(f)=[d^2_{2,m}(-\beta(f))d^2_{2,m'}(-\beta(f))+(-1)^{m+m'}d^2_{2,-m}(-\beta(f))d^2_{2,-m'}(-\beta(f))]\times\cos[(m-m')\alpha(f)]|h_\mathrm{D}(f)|^2, \tag{19}$$

$$q^\mathrm{sin}_{m,m'}(f)=[d^2_{2,m}(-\beta(f))d^2_{2,m'}(-\beta(f))+(-1)^{m+m'}d^2_{2,-m}(-\beta(f))d^2_{2,-m'}(-\beta(f))]\times\sin[(m-m')\alpha(f)]|h_\mathrm{D}(f)|^2\times(m,m'=-2,-1,0,1,2). \tag{20}$$

Since $q^\mathrm{cos}_{m,m'}=q^\mathrm{cos}_{m',m}=q^\mathrm{cos}_{-m,-m'}$ and $q^\mathrm{sin}_{m,m'}=-q^\mathrm{sin}_{m',m}=-q^\mathrm{sin}_{-m,-m'}$, only 15 of them are linearly independent. Thus, basis vectors are constructed for the 15 base waveforms. Those base waveforms are parametrized by the same parameters as those parametrizing $\{l_m(f)\}^2_{m=-2}$.

For IMRPhenomPv2 or IMRPhenomPv2_NRTidalv2 basis, we start with a training set of $\mathcal{O}(10^4)$ of waveforms. They are generated with $\mathcal{O}(10^3)$ random source parameters, for each of which there are five base waveforms for linear basis and 15 for quadratic basis. Validation is carried out with 10^5–10^6 random source parameters for IMRPhenomPv2 and 10^6–10^7 random source parameters for IMRPhenomPv2_NRTidalv2, where half of them incorporate random detector calibration errors. The base waveforms are normalized so that their aligned-spin limits have norm of unity:

$$\sum_{b=1}^{B}\frac{1}{T^{(b)}}\sum_{k}w^{(b)}(f_k^{(b)})|h_\mathrm{D}(f_k^{(b)})|^2=1, \tag{21}$$

$$\sum_{b=1}^{B}\frac{1}{\hat{T}^{(b)}}\sum_{k}w^{(b)}(\hat{f}_k^{(b)})|h_\mathrm{D}(\hat{f}_k^{(b)})|^2=1. \tag{22}$$

For the BNS IMRPhenomPv2 bases, the error tolerances for projection errors are 10^{-11} for all the linear bases and 3.6×10^{-16}, 4.4×10^{-16}, 5.1×10^{-16}, and 1.0×10^{-15} for quadratic bases with $T=512$, 256, 128, and 64 s, respectively. For the BNS IMRPhenomPv2_NRTidalv2 bases, they are 10^{-12} for all the linear bases and 3.6×10^{-17}, 4.4×10^{-17}, 5.1×10^{-17}, and 1.0×10^{-16} for quadratic bases with $T=512$, 256, 128, and 64 s, respectively. The lower tolerance values for IMRPhenomPv2_NRTidalv2 is to avoid any systematic biases from the ROQ approximation on the measurement of tidal effects. For the NSBH IMRPhenomPv2 bases, the error tolerances for projection errors are 10^{-10} for all the linear bases and 1.4×10^{-14}, 1.7×10^{-14}, 2.0×10^{-14}, 6.2×10^{-15}, and 1.2×10^{-14} for quadratic bases with $T=128$, 64, 32, 16, and 8 s, respectively.

E. Basis sizes and speedup gains

The sizes and speedup gains of the bases are presented in Table II. The low-spin IMRPhenomD bases are most compact and have only ∼100 basis elements for each in total. The IMRPhenomPv2 and IMRPhenomPv2_NRTidalv2 bases of the BNS mass range have several hundreds to ∼1000 basis elements for each, and the latter ones are larger due to the presence of tidal effects. The IMRPhenomPv2 bases of the intermediate mass range have higher basis sizes (up to several thousands) than the BNS bases due to the extended range of mass ratio. The sizes of the IMRPhenomXPHM bases range from several hundreds to thousands and significantly depend on chirp-mass range.

TABLE II. Chirp-mass partitions, frequency range, number of bases for each partition, basis sizes, and speedup gains of our ROQ bases. See Sec. III B for how the chirp-mass partitions are determined and Sec. III C for how each partition is divided into chirp-mass subdomains, for each of which ROQ basis is constructed. The speedup gains are measured speedup gains in likelihood evaluations. See Sec. III E for how they are measured.

	$\mathcal{M}(M_\odot)$		Frequencies (Hz)			Number of bases		Basis size		
Waveform	Min	Max	$f_{\rm low}$	$f_{\rm high}$	$1/T$	Linear	Quadratic	Linear	Quadratic	Speedup
IMRPhenomD	0.6	1.1	20	1024	1/512	149	1	126–137	24	250–460
	0.92	1.7	20	1024	1/256	74	1	120–130	25	110–210
	1.4	2.6	20	1024	1/128	37	1	112–122	28	58–100
	2.1	4.0	20	1024	1/64	20	1	109–117	32	29–43
IMRPhenomPv2 (BNS)	0.6	1.1	20	4096	1/512	15	15	639–788	454–646	790–1100
	0.92	1.7	20	4096	1/256	8	8	567–633	380–491	460–550
	1.4	2.6	20	4096	1/128	4	4	555–567	335–392	260–300
	2.1	4.0	20	2048	1/64	2	2	526–527	291–308	69–70
IMRPhenomPv2_NRTidalv2	0.6	1.1	20	4096	1/512	15	15	861–964	543–741	830–990
	0.92	1.7	20	4096	1/256	8	8	803–859	487–587	450–550
	1.4	2.6	20	4096	1/128	4	4	769–813	466–508	230–280
	2.1	4.0	20	2048	1/64	2	2	756–765	457–466	58–60
IMRPhenomPv2 (Intermediate)	1.4	2.6	20	1024	1/128	37	37	1540–1824	2808–3484	26–31
	2.1	4.0	20	1024	1/64	20	20	1716–2092	3256–4034	11–14
	3.3	6.3	20	1024	1/32	10	10	1979–2377	3840–4170	5.7–6.4
	5.2	11.0	20	1024	1/16	5	5	2304–2404	3794–4056	3.0–3.2
	8.7	21.0	20	1024	1/8	3	3	2062–2232	2525–3204	2.0–2.3
IMRPhenomXPHM	10.02	19.05	20	4096	1/32	23	23	1870–2092	2444–2619	14–19
	15.52	31.85	20	4096	1/16	15	15	1886–3095	2463–3095	4.9–8.0
	26.54	62.86	20	4096	1/8	11	11	1222–1836	1222–1836	2.4–3.8
	52.38	200.0	20	4096	1/4	11	11	308–806	308–806	1.9–3.7

The speedup gains presented in the tables are measured speedup gains in likelihood evaluations. For each chirp-mass subdomain, log-likelihood ratio is evaluated for 1000 random source parameter samples with and without the ROQ approximation and a speedup gain is calculated as the ratio of evaluation time for each sample. Each row in the table presents the range of the medians of measured speedup gains. For the low-spin IMRPhenomD bases, the measured speedup gains are ∼10 times lower than the expected speedup gains $K/(N_L + N_Q)$. It arises because waveform evaluations with those bases are so quick that other fixed costs such as precomputations of IMRPhenomD amplitude and phase coefficients dominate the cost. For the other cases, the measured speedup gains are roughly the same as the expected gains.

F. Likelihood errors

Figure 2 presents relative log-likelihood-ratio errors introduced by our ROQ bases. The errors have been computed for hundreds of source parameter samples per subdomain. For each parameter sample θ, a simulated signal is generated and considered as observed data $d_i(f) = h_i(f;\theta)$, and the relative error between $\ln\Lambda(\{d_i\}_{i=1}^{N_{\rm det}}|\theta)$ and $\ln\Lambda_{\rm ROQ}(\{d_i\}_{i=1}^{N_{\rm det}}|\theta)$ is computed. We consider only a single LIGO detector with its design sensitivity for this study. The source parameters contain amplitude and phase calibration errors at ten log-uniformly distributed frequency nodes, and interpolated calibration errors are multiplied by simulated signal and template waveform used for likelihood evaluations. Their values at the nodes are drawn from uniform distribution from −20% to 20% for amplitude and −15° to 15° for phase.

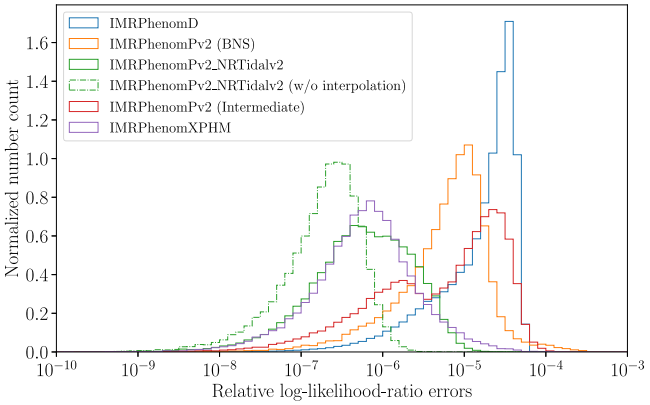

FIG. 2. Relative log-likelihood-ratio errors of our ROQ bases for random source parameter samples. They have been computed for hundreds of source parameter samples per basis subdomain.

As seen in the figure, the relative errors are $\lesssim 10^{-5}$ for the IMRPhenomPv2_NRTidalv2 bases and $\lesssim 10^{-4}$ for the other bases. The lower errors for IMRPhenomPv2_NRTidalv2 are due to the tighter error tolerances explained in the previous subsection. Since log-likelihood ratio is in the order of squared signal-to-noise ratio (SNR), the absolute log-likelihood-ratio errors are smaller than unity for SNR < 100, and our ROQ bases will not introduce biases in inference for typical SNR values observed by LVK detectors. We also note that the dominant errors of IMRPhenomPv2_NRTidalv2 do not come from the bases themselves but the numerical interpolation of ROQ weights over the coalescence time t_c. Relative errors computed when weights are calculated exactly are shown as the dashed-dotted line in the figure, and the errors get reduced to $\lesssim 10^{-6}$.

IV. APPLICATIONS

In this section, we demonstrate the usefulness of our rapid parameter estimation framework in various applications.

A. Rapid localization of BNS

Our rapid parameter estimation framework can be applied to rapid and accurate source localization of BNS signals for use in searches for their electromagnetic counterparts. In the current LVK alert system, the rapid source localization software, BAYESTAR [80,81], is run once CBC signal is detected. It utilizes output from a CBC search pipeline performing the matched-filtering [82–85] process on strain data. Specifically, it reads in matched-filter SNR time series for each detector and calculates the posterior probability distribution over sky location and luminosity distance to the source with a run-time of seconds. The input matched-filter SNR time series is computed with the best-matching template included in *a template bank*, a collection of simulated gravitational waveforms for various mass and spin values over which matched-filtering process is performed. To mitigate potential bias or loss of precision due to the mismatch between the signal and the best-matching template, BILBY is run to explore mass-spin space broader than that covered by the template bank and update the estimate of source location. Our low-spin IMRPhenomD ROQ bases have been developed specifically for speeding up this update procedure for BNS signal.

In this section, we benchmark the speed and localization accuracy of BILBY parameter estimation using the low-spin IMRPhenomD ROQ bases. We inject simulated signals, commonly called *injections*, into the third observing run (O3) data of the LIGO Hanford–LIGO Livingston–Virgo (HLV) detector network, which are publicly available [86]. Then, we recover their locations with our rapid parameter estimation framework and investigate the run-time and recovery accuracy. To benchmark its performance in a realistic situation, we perform end-to-end tests, where the simulated data are analyzed by a search pipeline and parameter estimation is run with the settings determined based on the search outputs. We also run BAYESTAR on the injections for comparison.

The source parameters of the injections are randomly drawn from the astrophysical distribution we assume. The detector-frame component masses are drawn from a uniform distribution across $1M_\odot \leq m_1, m_2 \leq 3M_\odot$. The spins are assumed to be parallel with the orbital angular momentum, and their components projected over the orbital angular momentum are drawn from a uniform distribution in $-0.05 \leq \chi_1, \chi_2 \leq 0.05$. The effects of tidal deformation of colliding bodies are not taken into account in injections. The source locations are distributed isotropically over sky location and uniformly in the cubic of luminosity distance D_L between 30 Mpc $\leq D_L \leq$ 600 Mpc. The distribution is isotropic in binary orientation and uniform over coalescence phase of binary motion. To avoid a lot of injections whose optimal SNRs are too small to detect, we preestimate the network optimal SNR, the root mean square of optimal SNRs at all the detectors, of each signal using reference O3 detector sensitivities, and inject only signals whose network optimal SNRs exceed $8\sqrt{2}$,[3] yielding 1047 injections in total. Those signals are synthesized based on the IMRPhenomD waveform model and injected into the O3 dataset between 13 June 2019 18:46 UTC and 16 August 2019 12:45 UTC. They are placed so that the interval of coalescence time t_c between neighboring injections is longer than 100 s to mitigate biases due to signal overlap.[4]

To simulate a gravitational-wave search, we use the GstLAL search pipeline (referred to as GstLAL hereafter) [87] with a template bank constructed based on a stochastic placement algorithm [88,89]. Each template is generated using the TaylorF2 waveform model [90]. For quick tests, we apply the matched-filter SNR maximized over coalescence phase and time as detection statistics rather than performing full likelihood analysis. Specifically, we apply the network matched-filter SNR above 12 and second largest SNR among the three detectors above 5.5 as the detection threshold, recovering 481 injections in total.

For 308 injections, data from all three detectors are available since all of the detectors were in observing mode. For the other 173 injections, one of the detectors was not in the observing mode. We refer to those two types of injections as *triple-detector injections* and *double-detector injections,* respectively. For each injection, the

[3]SNR of 8 is a typical threshold used to estimate the observable range of a detector, and SNRs exceeding 8 coincidentally at two detectors require the network SNR to be larger than $8\sqrt{2}$.

[4]We have found that only one pair of neighboring injections has time-frequency overlap, where the latter injection has lower masses than those of the earlier one and has longer signal duration. We expect its effects on our main results are negligible, since they are statistical results from hundreds of injections and the overlap has limited effects on the frequency integral of likelihood.

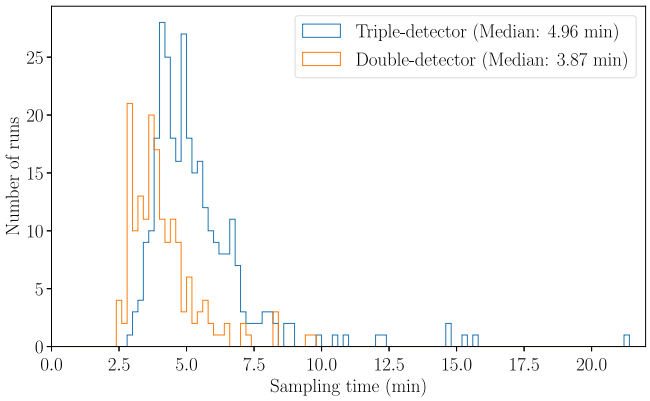

FIG. 3. Sampling time of BILBY with the low-spin IMRPhenomD ROQ bases for simulated BNS signals. The blue and orange histograms show sampling time for 304 triple-detector and 170 double-detector injections, respectively, and the median sampling time is presented in the legend.

detection criteria are typically met for multiple templates in the template bank. We refer to the template with the highest SNR among them as the *preferred template*, and the matched SNR time series computed with that template is used as input to BAYESTAR.

For BILBY analysis, we employ the DYNESTY sampler [91] with the acceptance-walk Markov chain Monte Carlo (MCMC) method. The number of live points and the average number of accepted MCMC jumps are set to 500 and 10, respectively. The sampling is parallelized with 24 processes. For each simulated signal, two independent runs are performed and their samples are combined. The PSD produced by GstLAL is used for likelihood evaluations.

The prior probability distribution is the same as that used to populate the injections, except for the explored range of $\mathcal{M}$. Since the chirp mass of the preferred template $\mathcal{M}_{\text{template}}$ is typically very close to its true value for BNS [92], its explored range is set to $0.995\mathcal{M}_{\text{template}} \leq \mathcal{M} \leq 1.005\mathcal{M}_{\text{template}}$. The explored range determined in that way includes the true chirp mass value except for three double-detector and four triple-detector injections. The errors of chirp-mass recoveries for those seven injections are higher than 20% and much higher than the errors for the other detections. Hence, we anticipate it is due to nonstationary or non-Gaussian noise around the injections. We exclude those seven injections in the main results presented in this section, while the effects of their inclusion are briefly explained in the text. We analytically marginalize over the coalescence phase, we marginalize over the luminosity distance using the look-up table method, and they are recovered at the postprocessing stage. Detector calibration uncertainties are not taken into account in these simulations and are not marginalized over.

Figure 3 presents the histogram of time taken by BILBY sampling for each signal. As seen in the figure, most of the runs complete within several minutes. The median run-time

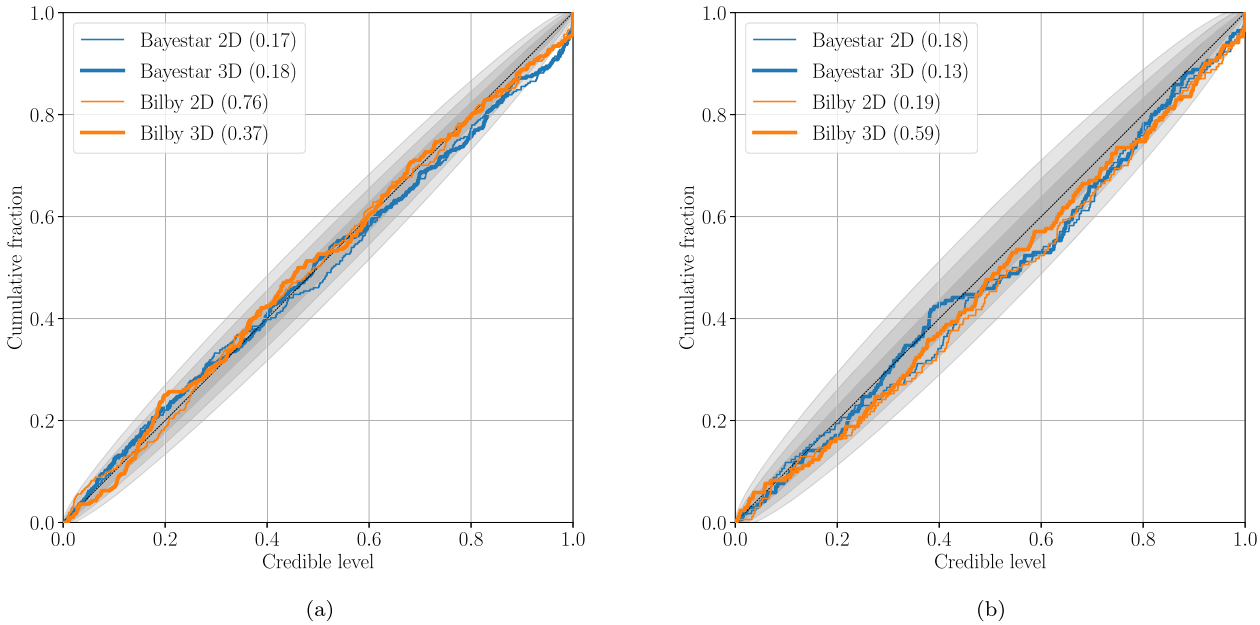

FIG. 4. *P-P* plots of estimates on sky location (2D, thin line) and three-dimensional location including distance (3D, thick line) from BAYESTAR (blue) and BILBY (orange) running on simulated BNS signals. The left and right panels show results from 304 triple-detector and 170 double-detector injections, respectively. The gray bands represent the 1, 2, and 3σ quantiles of statistical errors due to the finite number of samples. The p value of KS test between the observed credible levels and a uniform distribution for each case is presented in the legend. (a) triple-detector injections. (b) double-detector injections.

is 4.96 min for the triple-detector case and 3.87 min for the double-detector case. The runs are faster for the double-detector case since fewer operations are required to evaluate the likelihood due to fewer detectors. The runs are performed with an Intel Xeon Gold 6136 CPU with a clock rate of 3.0 GHz. In addition to the sampling time, precomputations of ROQ weights take ∼1 min for the triple-detector case and less for the double-detector case. Thus, the total run-time is minutes to 10 min, which provides enough time for follow-up observations of optical radiation fading away with the timescale of days [1].

Figure 4 presents *P-P* plots of sky location (2D) and three-dimensional location including distance (3D) for BAYESTAR and BILBY. They are visual tools to check whether the true signal parameters are found within the $X\%$ credible region $X\%$ of the time; i.e., it tests whether the posteriors have the correct statistical properties. If this is the case, the cumulative distribution of observed credible levels should be a diagonal line with statistical errors due to a finite number of samples. The gray bands represent the 1, 2, and 3σ quantiles of statistical errors. For the triple-detector case, both BAYESTAR and BILBY *P-P* plots are within the error band for credible level of $\lesssim 0.9$, while for the larger credible level BAYESTAR *P-P* plots go outside the band and BILBY performs better. The p value of KS test between the observed credible levels and a uniform distribution for each case is presented in the legend, and, in all the cases, the p values of BILBY are larger than those from BAYESTAR, implying BILBY produces more accurate results mitigating search biases. If the seven injections whose chirp-mass recoveries at the detection stage are bad are included, the p values for BAYESTAR 2D, BAYESTAR 3D, BILBY 2D, and BILBY 3D are degraded to (0.06, 0.07, 0.41, 0.42) for the triple-detector case and (0.09, 0.06, 0.10, 0.31) for the double-detector case.

Figure 5 presents cumulative distribution of searched areas and volumes, the areas and volumes one needs to search over from the most probable location to the least until reaching the true source location. As seen in the figure, searched areas and volumes are systematically smaller for BILBY, demonstrating that BILBY can reduce the area and volume follow-up observers need to search over.

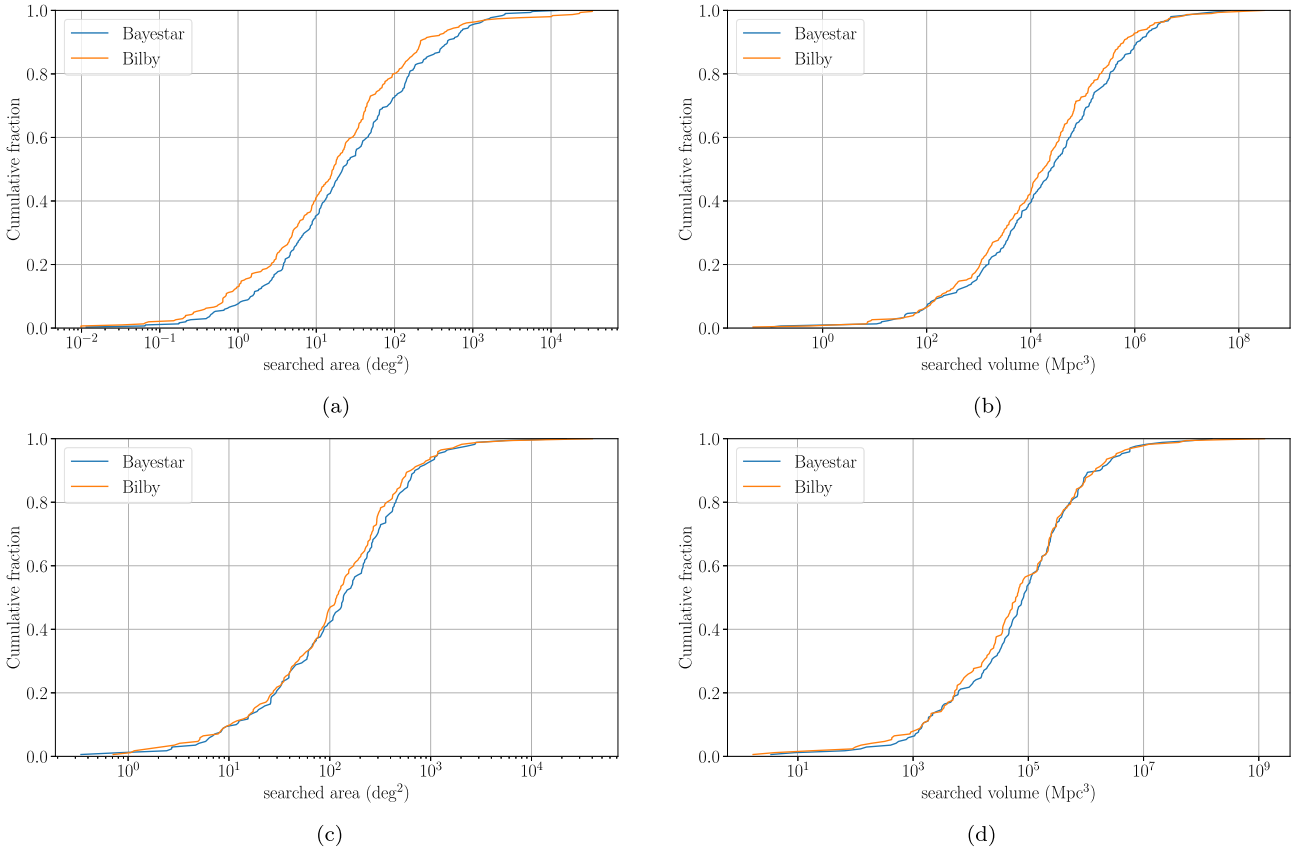

FIG. 5. Cumulative distribution of searched areas and volumes from BAYESTAR (blue) and BILBY (orange) running on simulated BNS signals. The top and bottom panels show searched areas and volumes respectively from 304 triple-detector (left) and 170 double-detector (right) injections. By updating BAYESTAR results with BILBY results, the median searched area (volume) is reduced from 21.8 deg^2 (2.47×10^4 Mpc^3) to 16.6 deg^2 (1.75×10^4 Mpc^3) for the triple-detector case and 137 deg^2 (8.30×10^4 Mpc^3) to 117 deg^2 (6.27×10^4 Mpc^3) for the double-detector case. (a) searched areas for triple-detector injections, (b) searched volumes for triple-detector injections, (c) searched areas for double-detector injections, and (d) searched volumes for double-detector injections.

The median searched area (volume) is reduced from 21.8 deg^2 (2.47×10^4 Mpc^3) to 16.6 deg^2 (1.75×10^4 Mpc^3) for the triple-detector case and 137 deg^2 (8.30×10^4 Mpc^3) to 117 deg^2 (6.27×10^4 Mpc^3) for the double-detector case. In summary, BILBY with our rapid parameter estimation technique improves the estimate on source location with the timescale of minutes.

B. Full BNS parameter estimation

We also demonstrate that our rapid parameter estimation framework can be applied to full BNS parameter estimation incorporating precession effects, tidal effects, and detector calibration uncertainties. We inject 100 simulated BNS signals into simulated Gaussian noise data of a HLV network drawn with their design sensitivities. We then recover their source parameters using BILBY with our rapid parameter estimation framework and investigate its accuracy using the *P-P* plots.

The injections are synthesized based on the IMRPhenomPv2_NRTidalv2 waveform model and analyzed with our IMRPhenomPv2_NRTidalv2 ROQ bases of the 256-s partition. The detector-frame component masses are uniformly distributed with the chirp-mass constraint $1.15M_\odot \leq \mathcal{M} \leq 1.25M_\odot$ and the mass ratio constraint $q \geq 0.125$. The spin magnitudes are uniformly distributed across $0 \leq a_1, a_2 \leq 0.4$, and the spin directions are isotropically distributed. The dimensionless tidal deformability values are uniformly distributed in $0 \leq \Lambda_1, \Lambda_2 \leq 5000$. The source locations are distributed uniformly in comoving volume and source frame time in the range 1 Mpc $\leq D_\mathrm{L} \leq 100$ Mpc. The distribution is isotropic in binary orientation, uniform over coalescence phase, and uniform over coalescence time within a time window with the width of 0.2 s. The median network optimal SNR of injections is 21.

Injections are multiplied by randomly generated calibration errors. The amplitude and phase errors are calculated as cubic spline interpolation of their values at ten frequency nodes log-uniformly distributed from 20 to 4096 Hz. We use calibration uncertainty budget of amplitude and phase at the GPS time of 1244415456 (12 June 2019, 22:57:18 UTC), which is publicly available [93] and presented in Fig. 6. The amplitude and phase errors at the nodes are drawn from Gaussian distribution whose mean matches the median shown as the dashed line and standard deviation is the half width of the 1σ band shown as the shaded region.

BILBY is run with the same prior probability distribution of source and calibration parameters as that to populate injections and the same PSD as that used for simulating Gaussian noise. The DYNESTY sampler is employed with the acceptance-walk MCMC method, 500 live points, and an average of 60 accepted MCMC jumps. We analytically marginalize over the coalescence phase, we marginalize over the luminosity distance using the look-up table method, and they are recovered at the postprocessing stage. The sampling is parallelized with 24 processes, and the median sampling time is 108 min. Without the ROQ approximation, the expected sampling time is on the order of a month.

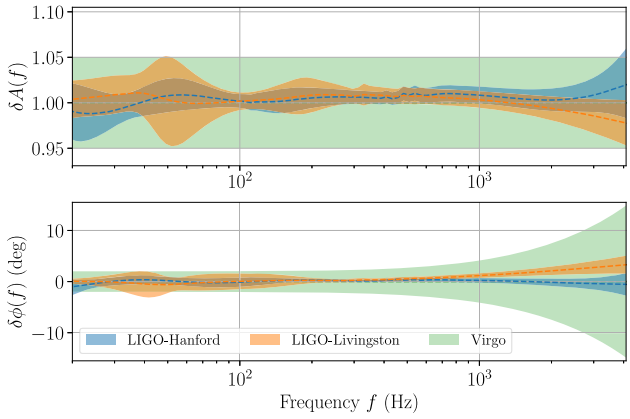

FIG. 6. Detector calibration uncertainties of amplitude (top) and phase (bottom) used for simulations. The dashed lines represent the median values, and the shaded regions represent the 1σ uncertainties.

Figure 7 presents *P-P* plots of all the source parameters. As seen in the figure, they are well within the error band,

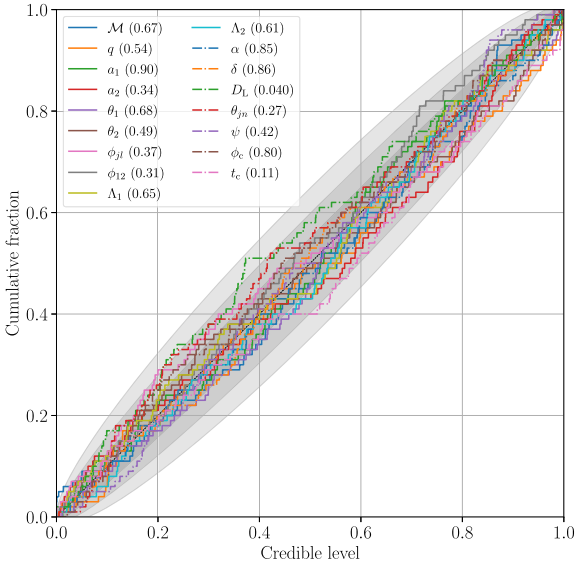

FIG. 7. *P-P* plots of all the parameters recovered by BILBY with the IMRPhenomPv2_NRTidalv2 ROQ bases running on 100 injections. The recovered parameters are detector-frame chirp mass ($\mathcal{M}$), mass ratio (q), spin magnitudes (a_1 and a_2), spin angles (θ_1, θ_2, ϕ_{jl}, and ϕ_{12}), dimensionless tidal deformability parameters (Λ_1 and Λ_2), right ascension (α) and declination (δ) of the source location, luminosity distance to the source (D_L), inclination angle between total angular momentum and line of sight (θ_{jn}), polarization angle (ψ), coalescence phase (ϕ_c), and coalescence time (t_c). The gray bands represent the 1, 2, and 3σ quantiles of statistical errors due to the finite number of samples. The p value of KS test between the observed credible levels and a uniform distribution from 0 to 1 for each parameter is presented in the legend. The combined p value, the probability that each p value is drawn from a uniform distribution from 0 to 1, is 0.78.

and p values in the legend imply the observed deviations are consistent with statical errors. The combined p value, the probability that each p value is drawn from a uniform distribution from 0 to 1, is 0.78. Hence, we conclude our rapid parameter estimation framework can also be applied to full and accurate parameter estimation analysis of BNS signal.

C. BBH rapid parameter estimation

Finally, we also demonstrate the application of the IMRPhenomXPHM bases to BBH signals. We inject 100 simulated BBH signals synthesized based on the IMRPhenomXPHM waveform model into simulated Gaussian noise of a HLV network and analyze them with our IMRPhenomXPHM 8-s ROQ bases.

The distribution to populate injections and analysis settings are almost same as what are used in the previous section with the following differences. The mass and spin range considered here is $26.54M_\odot \leq \mathcal{M} \leq 62.86M_\odot$, $1/20 \leq q \leq 1$, and $0 \leq a_1, a_2 \leq 0.99$, and tidal deformability values are fixed to zero. The distance range is 10 Mpc $\leq D_{\rm L} \leq$ 1000 Mpc, yielding the median network optimal SNR value of 30. Only the luminosity distance parameter is marginalized over with the look-up table method, since analytical phase marginalization is not applicable when gravitational-wave higher multipole moments are present. The median sampling time is 198 min.

Figure 8 presents P-P plots of all the source parameters. The p values in the legend imply the observed deviations from the diagonal line is consistent with statistical errors, and the combined p value is 0.42. This demonstrates our rapid parameter estimation framework can also be applied to full parameter estimation analysis of BBH signal.

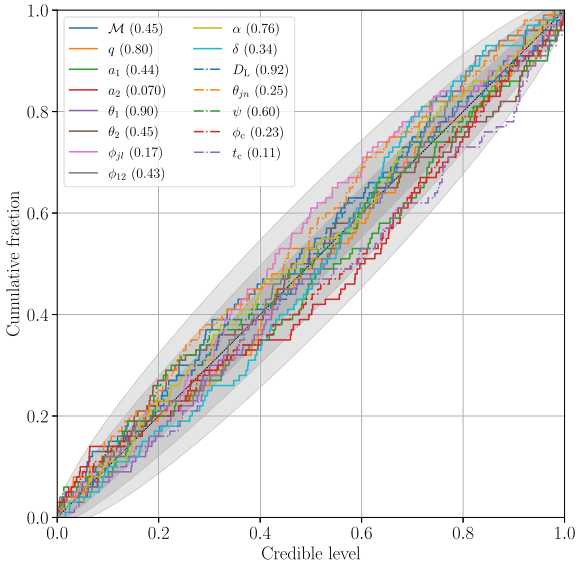

FIG. 8. P-P plots of all the parameters recovered by BILBY with the IMRPhenomXPHM ROQ bases running on 100 injections. The combined p value is 0.42. See the caption of Fig. 7 for the definition of parameter symbols and the combined p value and the meaning of numbers in the legend and gray bands in the plot.

V. CONCLUSION

In this paper, we have presented a rapid parameter estimation framework using multiple ROQ bases of state-of-the-art GW signal models describing a broad range of CBC sources. Each basis is constructed over narrow parameter space to produce a significant speedup, while the union of the bases covers broad parameter space. Hence, our framework can accelerate parameter estimation significantly without sacrificing accuracy. Based on this idea, we have developed sets of ROQ bases constructed over narrow chirp-mass subdomains. As demonstrated in Sec. IV, our framework and new ROQ bases enable improved source localization of BNS signal with the timescale of minutes, as well as more detailed parameter estimation taking into account general spin configurations, GW higher multipole moments, and tidal deformation of colliding objects, with the timescale of hours. The combined use cases greatly improves the scalability of parameter estimation workflows during observing runs, especially as event rates increase due to improved detector sensitivity. Our multiple-ROQ-bases framework has been implemented in one of the LVK parameter estimation engines, BILBY, and that framework as well as our newly developed ROQ bases are being employed by the automated parameter estimation analysis of the LVK O4 alert system, circulating source location estimates to followup observers [58–61].

One possible extension of this work is to extend the lower mass end of IMRPhenomXPHM bases. While constructing IMRPhenomXPHM bases in the lower mass region is computationally costly and consumes excessive amounts of memory, they are useful for rapid source localization of NSBH signal taking into account orbital precession and GW higher multipole moments. As we did for the other bases, directly constructing multibanded bases can significantly reduce the memory consumption. Dimensionality reduction by a sophisticated choice of base waveforms, as done for the construction of the IMRPhenomPv2 and IMRPhenomPv2_NRTidalv2 bases, may also be possible.

Another direction is to construct bases for other BNS waveform models, such as SEOBNRv4ROM_NRTidalv2 [72,94], IMRPhenomXP_NRTidalv2, and SEOBNRv4T_Surrogate [95]. Those bases enable us to infer tidal deformability parameters with multiple waveform models, mitigating waveform systematics to obtain accurate constraints on the nuclear equation of state.

ACKNOWLEDGMENTS

We are grateful to Carl-Johan Haster for careful reading of the manuscript and useful feedback. S. M. and R. S.

thank the Institute for Computational and Experimental Research in Mathematics (ICERM) at Brown University for the support during their Fall 2020 reunion event, where considerable progress was made on this work. S. M. is supported by JSPS Grant-in-Aid for Transformative Research Areas (A) No. 23H04891 and No. 23H04893. This research was supported in part by the Australian Research Council Centre of Excellence for Gravitational Wave Discovery (OzGrav), through Project No. CE170100004. S. S. is a recipient of an ARC Discovery Early Career Research Award (DE220100241). C. T. is supported by an MIT Kavli fellowship. A. Z. is supported by National Science Foundation (NSF) Grant No. PHY-2308833. This work has been assigned preprint numbers LIGO-P2300205 and UTWI-27-2023. The authors are grateful for computational resources provided by the LIGO Laboratory and the Leonard E Parker Center for Gravitation, Cosmology and Astrophysics at the University of Wisconsin–Milwaukee supported by National Science Foundation Grants No. PHY-1626190, No. PHY-1700765, No. PHY-0757058 and No. PHY-0823459. This research has made use of data or software obtained from the Gravitational Wave Open Science Center, a service of the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation, 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 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, and Spain. K. A. G. R. A. is supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; and Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan.

APPENDIX A: MULTIBAND DECOMPOSITION OF ROQ LIKELIHOOD RATIO

In this appendix, we obtain a multibanded form of ROQ likelihood ratio. In the multiband approximation developed in [34], the total frequency range is divided into B overlapping frequency bands $f_{\rm s}^{(b)} \le f \le f_{\rm e}^{(b)}$ $(b = 1, 2, \ldots, B)$ with a set of smooth window functions $\{w^{(b)}(f)\}_{b=1}^{B}$. The bth frequency band is constructed so that signal duration from the starting frequency $f_{\rm s}^{(b)}$ is smaller than a certain duration value $T^{(b)}$, where $T = T^{(1)} > T^{(2)} > \cdots > T^{(B)}$. The inner products $(d_i, h_i(\theta))_i$ and $(h_i(\theta), h_i(\theta))_i$ are then approximated into the following forms [see Eqs. (24) and (45) in [34]]:

$$(d_i, h_i(\theta))_i \simeq \sum_{b=1}^{B} \frac{4}{T^{(b)}} \mathrm{Re}\left[\sum_{k=\lceil f_{\rm s}^{(b)} T^{(b)} \rceil}^{\lfloor f_{\rm e}^{(b)} T^{(b)} \rfloor} w^{(b)}(f_k^{(b)}) \tilde{D}_k^{(b)*} h(f_k^{(b)}) \right], \tag{A1}$$

$$(h_i(\theta), h_i(\theta))_i \simeq \sum_{b=1}^{B} \frac{4}{\hat{T}^{(b)}} \sum_{k=\lceil f_{\rm s}^{(b)} \hat{T}^{(b)} \rceil}^{\lfloor f_{\rm e}^{(b)} \hat{T}^{(b)} \rfloor} w^{(b)}(\hat{f}_k^{(b)}) \tilde{I}_{{\rm c},k}^{(b)} |h(\hat{f}_k^{(b)})|^2, \tag{A2}$$

where $f_k^{(b)} = k/T^{(b)}$, $\tilde{D}_k^{(b)*}$ is a quantity dependent on data and PSD, $\hat{T}^{(b)} = \min[2T^{(b)}, T]$, $\hat{f}_k^{(b)} = k/\hat{T}^{(b)}$, and $\tilde{I}_{{\rm c},k}^{(b)}$ is a quantity dependent on PSD.

Substituting Eqs. (6) and (7) into Eqs. (A1) and (A2) and substituting the approximate inner products into the log-likelihood ratio (5), we obtain

$$\ln \Lambda_{\rm ROQ}^{\rm MB} = \sum_{i=1}^{N_{\rm det}} \left[L_i^{\rm MB}(\theta) - \frac{1}{2} Q_i^{\rm MB}(\theta) \right], \tag{A3}$$

$$L_i^{\rm MB}(\theta) = \mathrm{Re}\left[\sum_{I=1}^{N_L} h_i(F_I; \theta) \omega_{I,i}^{\rm MB}(t_c) \right], \tag{A4}$$

$$Q_i^{\rm MB}(\theta) = \sum_{J=1}^{N_Q} |h_i(\mathcal{F}_J; \theta)|^2 \psi_{J,i}^{\rm MB}, \tag{A5}$$

where

$$\omega_{I,i}^{\rm MB}(t_{\rm c}) = \sum_{b=1}^{B} \frac{4}{T^{(b)}} \mathrm{Re}\left[\sum_{k=\lceil f_{\rm s}^{(b)} T^{(b)} \rceil}^{\lfloor f_{\rm e}^{(b)} T^{(b)} \rfloor} w^{(b)}(f_k^{(b)}) \tilde{D}_k^{(b)*} B_I(f_k^{(b)}) e^{-2\pi i f_k^{(b)} t_c} \right], \tag{A6}$$

$$\psi_I^{\rm MB} = \sum_{b=1}^{B} \frac{4}{\hat{T}^{(b)}} \sum_{k=\lceil f_{\rm s}^{(b)} \hat{T}^{(b)} \rceil}^{\lfloor f_{\rm e}^{(b)} \hat{T}^{(b)} \rfloor} w^{(b)}(\hat{f}_k^{(b)}) \tilde{I}_{c,k}^{(b)} C_J(\hat{f}_k^{(b)}). \tag{A7}$$

The integration weights $\omega_{I,i}^{\rm MB}(t_{\rm c})$ and $\psi_{J,i}^{\rm MB}$ can be computed with the multibanded ROQ bases $\{\{B_I(f_k^{(b)})\}_k\}_{b=1}^{B}$ and $\{\{C_J(\hat{f}_k^{(b)})\}_k\}_{b=1}^{B}$, and, hence, only those downsampled components need to be stored.

APPENDIX B: CUSPS IN IMRPhenomPv2 WAVEFORMS

IMRPhenomPv2 waveforms have cusps in a certain mass-spin space, where mass ratio q is relatively low ($q \lesssim 0.4$) and total spin angular momentum projected onto the orbital angular momentum is negative. The cusps come from the Wigner coefficients $d^2_{2,m}(-\beta(f))$, where $\beta(f)$ is the opening angle between the total angular momentum and the orbital angular momentum at a GW frequency f.

By the definition of $\beta(f)$, $\cos\beta(f)$ can be calculated as follows:

$$\cos\beta(f) = \pm(1 + (s(f))^2)^{-\frac{1}{2}}, \tag{B1}$$

where $s(f) = S_\perp/(L(f) + S_\parallel)$, $L(f)$ is the norm of the orbital angular momentum at a GW frequency f and $S_\parallel$ and $S_\perp$ are the components of total spin angular momentum parallel with and orthogonal to the orbital angular momentum, respectively. Mathematically, the sign of $\cos\beta(f)$ should follow the sign of $L(f) + S_\parallel$. However, in IMRPhenomPv2 the positive sign is always taken regardless of the sign of $L(f) + S_\parallel$. Thus, $\cos\beta(f)$ has a cusp at a frequency where $L(f) + S_\parallel$ crosses 0, and, hence, $d^2_{2,m}(-\beta(f))$ also has a cusp there, since it depends on $\cos\beta(f)$, as shown in Fig. 9.

In the parameter space we consider, $L(f) + S_\parallel$ is always positive at $f = f_{\rm low} = 20$ Hz. Hence, the necessary and sufficient condition for the existence of a waveform cusp is that the minimum of $L(f) + S_\parallel$ below $f = f_{\rm high}$ is negative. In IMRPhenomPv2, $L(f)$ is calculated with the nonspinning second-order post-Newtonian formula:

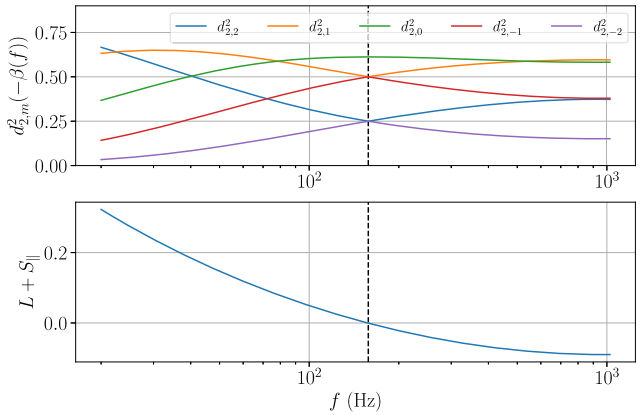

FIG. 9. Wigner coefficients $d^2_{2,m}(-\beta(f))$ (top) and the total angular momentum projected onto the orbital angular momentum (bottom) of IMRPhenomPv2 for $m_1 = 8M_\odot$, $m_2 = 1M_\odot$, $\chi_1 = -0.5$, $\chi_2 = -0.5$ and, $\chi_{\rm p} = 0.5$. The dashed vertical line indicates the frequency where the projected total angular momentum crosses 0.

$$L(f) = \frac{\eta(m_1 + m_2)^2}{v} \times \left(1 + \left(\frac{3}{2} + \frac{\eta}{6}\right)v^2 + \left(\frac{27}{8} - \frac{19\eta}{8} - \frac{\eta^2}{24}\right)v^4\right), \tag{B2}$$

where $\eta = q/(1+q)^2$ and $v = (\pi(m_1 + m_2)f)^{\frac{1}{3}}$. Within the frequency range from $f_{\rm low}$ to $f_{\rm high}$, it gets its minimum at $v = \hat{v}$, where

$$\hat{v} = \min\left[\sqrt{\frac{2(9 + \eta - \sqrt{1539 - 1008\eta - 17\eta^2})}{3(-81 + 57\eta + \eta^2)}}, (\pi(m_1 + m_2)f_{\rm high})^{\frac{1}{3}}\right]. \tag{B3}$$

On the other hand, $S_\parallel = m_1^2\chi_1 + m_2^2\chi_2$. Thus, the mass-spin region where waveform has a cusp is expressed by

$$\frac{\eta}{\hat{v}}\left(1 + \left(\frac{3}{2} + \frac{\eta}{6}\right)\hat{v}^2 + \left(\frac{27}{8} - \frac{19\eta}{8} - \frac{\eta^2}{24}\right)\hat{v}^4\right) + \frac{\chi_1 + q^2\chi_2}{(1+q)^2} < 0. \tag{B4}$$

APPENDIX C: BASE WAVEFORMS FOR IMRPhenomPv2

GW polarizations of IMRPhenomPv2 are given by

$$h_+(f) = \frac{1}{2}e^{-2i\epsilon(f)}(T(f) + T_c(f))h_{\rm D}(f), \tag{C1}$$

$$h_\times(f) = \frac{i}{2}e^{-2i\epsilon(f)}(T(f) - T_c(f))h_{\rm D}(f), \tag{C2}$$

where

$$T(f) = \sum_{m=-2}^{2}(-1)^m e^{im\alpha(f)}d^2_{2,m}(-\beta(f))Y^2_{2,-m}(\theta_J, 0), \tag{C3}$$

$$T_c(f) = \sum_{m=-2}^{2}e^{im\alpha(f)}d^2_{2,m}(-\beta(f))Y^2_{2,m}(\theta_J, 0), \tag{C4}$$

and $Y^s_{l,m}(\theta,\phi)$ is spin-weighted spherical harmonics. Hence, it is evident that $h_+(f)$ and $h_\times(f)$ are linear combinations of $l_m(f)$ $(m = -2, -1, 0, 1, 2)$ given by Eq. (18).

On the other hand, $|F_+h_+ + F_\times h_\times|^2$ contains the following products:

$$|h_{+/\times}(f)|^2 = \frac{1}{4}(|T(f)|^2 + |T_c(f)|^2 \pm 2\mathrm{Re}[T^*(f)T_c(f)])|h_{\mathrm{D}}(f)|^2, \quad \text{(C5)}$$

$$\mathrm{Re}[h_+^*(f)h_\times(f)] = \frac{1}{2}\mathrm{Im}[T^*(f)T_c(f)]|h_{\mathrm{D}}(f)|^2. \quad \text{(C6)}$$

$|T(f)|^2 + |T_c(f)|^2$ is linear combination of $q^{\cos}_{m,m'}(f)$ as shown below:

$$\begin{aligned}
&|T(f)|^2 + |T_c(f)|^2 \\
&= \sum_{m,m'}[(-1)^{m+m'}Y_{-m}Y_{-m'} + Y_mY_{m'}] \\
&\quad \times d^2_{2,m}(-\beta(f))d^2_{2,m'}(-\beta(f))\mathrm{e}^{\mathrm{i}(m-m')\alpha(f)} \\
&= \sum_{m,m'}[(-1)^{m+m'}Y_{-m}Y_{-m'} + Y_mY_{m'}] \\
&\quad \times d^2_{2,m}(-\beta(f))d^2_{2,m'}(-\beta(f))\cos[(m-m')\alpha(f)] \\
&= \frac{1}{2}\sum_{m,m'}[(-1)^{m+m'}Y_{-m}Y_{-m'} + Y_mY_{m'}]q^{\cos}_{m,m'}(f), \quad \text{(C7)}
\end{aligned}$$

where Y_m represents $Y^2_{2,m}(\theta_J, 0)$. Similarly, $\mathrm{Re}[T^*(f)T_c(f)]$ is linear combination of $q^{\cos}_{m,m'}(f)$:

$$\begin{aligned}
&\mathrm{Re}[T^*(f)T_c(f)] \\
&= \sum_{m,m'}(-1)^{m'}Y_mY_{-m'} \\
&\quad \times d^2_{2,m}(-\beta(f))d^2_{2,m'}(-\beta(f))\cos[(m-m')\alpha(f)] \\
&= \frac{1}{2}\sum_{m,m'}[(-1)^{m'}Y_mY_{-m'} + (-1)^mY_{-m}Y_{m'}] \\
&\quad \times d^2_{2,m}(-\beta(f))d^2_{2,m'}(-\beta(f))\cos[(m-m')\alpha(f)] \\
&= \frac{1}{4}\sum_{m,m'}[(-1)^{m'}Y_mY_{-m'} + (-1)^mY_{-m}Y_{m'}]q^{\cos}_{m,m'}(f), \quad \text{(C8)}
\end{aligned}$$

and $\mathrm{Im}[T^*(f)T_c(f)]$ is linear combination of $q^{\sin}_{m,m'}(f)$:

$$\begin{aligned}
&\mathrm{Im}[T^*(f)T_c(f)] \\
&= \sum_{m,m'}(-1)^{m'}Y_mY_{-m'} \\
&\quad \times d^2_{2,m}(-\beta(f))d^2_{2,m'}(-\beta(f))\sin[(m-m')\alpha(f)] \\
&= \frac{1}{2}\sum_{m,m'}[(-1)^{m'}Y_mY_{-m'} - (-1)^mY_{-m}Y_{m'}] \\
&\quad \times d^2_{2,m}(-\beta(f))d^2_{2,m'}(-\beta(f))\sin[(m-m')\alpha(f)] \\
&= \frac{1}{4}\sum_{m,m'}[(-1)^{m'}Y_mY_{-m'} - (-1)^mY_{-m}Y_{m'}]q^{\sin}_{m,m'}(f). \quad \text{(C9)}
\end{aligned}$$

Hence, $|F_+h_+ + F_\times h_\times|^2$ is linear combination of $q^{\cos}_{m,m'}(f)$ and $q^{\sin}_{m,m'}(f)$.

[1] B. P. Abbott *et al.* (LIGO Scientific, Virgo, Fermi GBM, INTEGRAL, IceCube, AstroSat Cadmium Zinc Telluride Imager Team, IPN, Insight-Hxmt, ANTARES, Swift, AGILE Team, 1M2H Team, Dark Energy Camera GW-EM, DES, DLT40, GRAWITA, Fermi-LAT, ATCA, ASKAP, Las Cumbres Observatory Group, OzGrav, DWF (Deeper Wider Faster Program), AST3, CAASTRO, VINROUGE, MASTER, J-GEM, GROWTH, JAGWAR, CaltechNRAO, TTU-NRAO, NuSTAR, Pan-STARRS, MAXI Team, TZAC Consortium, KU, Nordic Optical Telescope, ePESSTO, GROND, Texas Tech University, SALT Group, TOROS, BOOTES, MWA, CALET, IKI-GW Follow-up, H.E.S.S., LOFAR, LWA, HAWC, Pierre Auger, ALMA, Euro VLBI Team, Pi of Sky, Chandra Team at McGill University, DFN, ATLAS Telescopes, High Time Resolution Universe Survey, RIMAS, RATIR, SKA South Africa/MeerKAT Collaborations), Astrophys. J. Lett. **848**, L12 (2017).

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

[3] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. X **9**, 011001 (2019).

[4] A. Goldstein *et al.*, Astrophys. J. Lett. **848**, L14 (2017).

[5] V. Savchenko *et al.*, Astrophys. J. Lett. **848**, L15 (2017).

[6] B. P. Abbott *et al.* (LIGO Scientific, Virgo, Fermi-GBM, and INTEGRAL Collaborations), Astrophys. J. Lett. **848**, L13 (2017).

[7] P. Cowperthwaite *et al.*, Astrophys. J. Lett. **848**, L17 (2017).

[8] P. Evans *et al.*, Science **358**, 1565 (2017).

[9] I. Arcavi *et al.*, Nature (London) **551**, 64 (2017).

[10] Y. Utsumi *et al.* (J-GEM Collaboration), Publ. Astron. Soc. Jpn. **69**, 101 (2017).

[11] K. Alexander *et al.*, Astrophys. J. Lett. **848**, L21 (2017).

[12] E. Troja *et al.*, Nature (London) **551**, 71 (2017).

[13] M. Drout *et al.*, Science **358**, 1570 (2017).

[14] M. Kasliwal *et al.*, Science **358**, 1559 (2017).

[15] N. Tanvir *et al.*, Astrophys. J. Lett. **848**, L27 (2017).

[16] E. Pian *et al.*, Nature (London) **551**, 67 (2017).

[17] M. Tanaka *et al.*, Publ. Astron. Soc. Jpn. **69**, 102 (2017).

[18] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Phys. Rev. Lett. **121**, 161101 (2018).

[19] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Classical Quantum Gravity **37**, 045006 (2020).

[20] B. P. Abbott *et al.* (LIGO Scientific, Virgo, 1M2H, Dark Energy Camera GW-E, DES, DLT40, Las Cumbres Observatory, VINROUGE, MASTER Collaborations), Nature (London) **551**, 85 (2017).
[21] K. Hotokezaka, E. Nakar, O. Gottlieb, S. Nissanke, K. Masuda, G. Hallinan, K. P. Mooley, and A. T. Deller, Nat. Astron. **3**, 940 (2019).
[22] P. Creminelli and F. Vernizzi, Phys. Rev. Lett. **119**, 251302 (2017).
[23] J. M. Ezquiaga and M. Zumalacárregui, Phys. Rev. Lett. **119**, 251304 (2017).
[24] T. Baker, E. Bellini, P. Ferreira, M. Lagos, J. Noller, and I. Sawicki, Phys. Rev. Lett. **119**, 251301 (2017).
[25] B. P. Abbott *et al.* (KAGRA, LIGO Scientific, and VIRGO Collaborations), Living Rev. Relativity **21**, 3 (2018).
[26] J. Aasi *et al.* (LIGO Scientific Collaboration), Classical Quantum Gravity **32**, 074001 (2015).
[27] F. Acernese *et al.* (VIRGO Collaboration), Classical Quantum Gravity **32**, 024001 (2015).
[28] T. Akutsu *et al.* (KAGRA Collaboration), Prog. Theor. Exp. Phys. **2021**, 05A102 (2021).
[29] P. Petrov, L. P. Singer, M. W. Coughlin, V. Kumar, M. Almualla, S. Anand, M. Bulla, T. Dietrich, F. Foucart, and N. Guessoum, Astrophys. J. **924**, 54 (2022).
[30] N. J. Cornish, arXiv:1007.4820.
[31] N. J. Cornish, Phys. Rev. D **103**, 104057 (2021).
[32] N. Leslie, L. Dai, and G. Pratten, Phys. Rev. D **104**, 123030 (2021).
[33] S. Vinciguerra, J. Veitch, and I. Mandel, Classical Quantum Gravity **34**, 115006 (2017).
[34] S. Morisaki, Phys. Rev. D **104**, 044062 (2021).
[35] L. Pathak, A. Reza, and A. S. Sengupta, Phys. Rev. D **108**, 064055 (2023).
[36] C. Pankow, P. Brady, E. Ochsner, and R. O'Shaughnessy, Phys. Rev. D **92**, 023002 (2015).
[37] J. Lange, R. O'Shaughnessy, and M. Rizzo, arXiv:1805.10457.
[38] D. Wysocki, R. O'Shaughnessy, J. Lange, and Y.-L. L. Fang, Phys. Rev. D **99**, 084026 (2019).
[39] J. Wofford *et al.*, Phys. Rev. D **107**, 024040 (2023).
[40] C. A. Rose, V. Valsan, P. R. Brady, S. Walsh, and C. Pankow, arXiv:2201.05263.
[41] R. J. E. Smith, G. Ashton, A. Vajpeyi, and C. Talbot, Mon. Not. R. Astron. Soc. **498**, 4492 (2020).
[42] C. Talbot, R. Smith, E. Thrane, and G. B. Poole, Phys. Rev. D **100**, 043030 (2019).
[43] S. R. Green, C. Simpson, and J. Gair, Phys. Rev. D **102**, 104057 (2020).
[44] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, Phys. Rev. Lett. **127**, 241103 (2021).
[45] M. Dax, S. R. Green, J. Gair, M. Pürrer, J. Wildberger, J. H. Macke, A. Buonanno, and B. Schölkopf, Phys. Rev. Lett. **130**, 171403 (2023).
[46] M. J. Williams, J. Veitch, and C. Messenger, Phys. Rev. D **103**, 103006 (2021).
[47] C. Chatterjee, L. Wen, D. Beveridge, F. Diakogiannis, and K. Vinsen, arXiv:2207.14522.
[48] E. Lee, S. Morisaki, and H. Tagoshi, Phys. Rev. D **105**, 124057 (2022).
[49] J. Roulet, S. Olsen, J. Mushkin, T. Islam, T. Venumadhav, B. Zackay, and M. Zaldarriaga, Phys. Rev. D **106**, 123015 (2022).
[50] T. Islam, J. Roulet, and T. Venumadhav, arXiv:2210.16278.
[51] K. W. K. Wong, M. Isi, and T. D. P. Edwards, Astrophys. J. **958**, 129 (2023).
[52] V. Tiwari, C. Hoy, S. Fairhurst, and D. MacLeod, Phys. Rev. D **108**, 023001 (2023).
[53] P. Canizares, S. E. Field, J. Gair, V. Raymond, R. Smith, and M. Tiglio, Phys. Rev. Lett. **114**, 071104 (2015).
[54] R. Smith, S. E. Field, K. Blackburn, C.-J. Haster, M. Pürrer, V. Raymond, and P. Schmidt, Phys. Rev. D **94**, 044031 (2016).
[55] S. Morisaki and V. Raymond, Phys. Rev. D **102**, 104020 (2020).
[56] G. Ashton *et al.*, Astrophys. J. Suppl. Ser. **241**, 27 (2019).
[57] I. M. Romero-Shaw *et al.*, Mon. Not. R. Astron. Soc. **499**, 3295 (2020).
[58] LIGO Scientific Collaboration, VIRGO Collaboration and Kagra Collaboration, GRB Coordinates Network **33816**, 1 (2023).
[59] LIGO Scientific Collaboration, VIRGO Collaboration, and Kagra Collaboration, GRB Coordinates Network **33891**, 1 (2023).
[60] LIGO Scientific Collaboration, VIRGO Collaboration, and Kagra Collaboration, GRB Coordinates Network **33919**, 1 (2023).
[61] Ligo Scientific Collaboration, VIRGO Collaboration, and Kagra Collaboration, GRB Coordinates Network **34087**, 1 (2023).
[62] P. Whittle, *Hypothesis Testing in Times Series Analysis* (Almqvist & Wiksells Boktryckeri AB, Oxford, England, Uppsala, 1951).
[63] R. Abbott *et al.* (LIGO Scientific, VIRGO, and KAGRA Collaborations), Phys. Rev. X **13**, 041039 (2023).
[64] R. Abbott *et al.* (LIGO Scientific and VIRGO Collaborations), Phys. Rev. D **108**, DS12984 (2023).
[65] T. M. Tauris *et al.*, Astrophys. J. **846**, 170 (2017).
[66] K.-W. Lo and L.-M. Lin, Astrophys. J. **728**, 12 (2011).
[67] M. Burgay *et al.*, Nature (London) **426**, 531 (2003).
[68] E. E. Flanagan and T. Hinderer, Phys. Rev. D **77**, 021502 (2008).
[69] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. Jiménez Forteza, and A. Bohé, Phys. Rev. D **93**, 044006 (2016).
[70] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. Jiménez Forteza, and A. Bohé, Phys. Rev. D **93**, 044007 (2016).
[71] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Phys. Rev. Lett. **113**, 151101 (2014).
[72] T. Dietrich, A. Samajdar, S. Khan, N. K. Johnson-McDaniel, R. Dudi, and W. Tichy, Phys. Rev. D **100**, 044003 (2019).
[73] G. Pratten *et al.*, Phys. Rev. D **103**, 104056 (2021).
[74] R. Abbott *et al.* (KAGRA, VIRGO, and LIGO Scientific Collaborations), Phys. Rev. X **13**, 011048 (2023).
[75] F. Foucart, Front. Astron. Space Sci. **7**, 46 (2020).
[76] J. Veitch *et al.*, Phys. Rev. D **91**, 042003 (2015).

[77] S. E. Field, C. R. Galley, F. Herrmann, J. S. Hesthaven, E. Ochsner, and M. Tiglio, Phys. Rev. Lett. **106**, 221102 (2011).
[78] S. Vitale, W. Del Pozzo, T. G. F. Li, C. Van Den Broeck, I. Mandel, B. Aylott, and J. Veitch, Phys. Rev. D **85**, 064034 (2012).
[79] P. Schmidt, F. Ohme, and M. Hannam, Phys. Rev. D **91**, 024043 (2015).
[80] L. P. Singer and L. R. Price, Phys. Rev. D **93**, 024013 (2016).
[81] L. P. Singer *et al.*, Astrophys. J. Lett. **829**, L15 (2016).
[82] L. Wainstein, V. D. Zubakov, and A. A. Mullin, *Extraction of Signals from Noise* (Dover Publications, New York, 1970).
[83] B. S. Sathyaprakash, Phys. Rev. D **44**, 3819 (1991).
[84] L. S. Finn, Phys. Rev. D **46**, 5236 (1992).
[85] L. S. Finn, Phys. Rev. D **47**, 2198 (1993).
[86] R. Abbott *et al.* (LIGO Scientific, VIRGO, and KAGRA Collaborations), Astrophys. J. Suppl. Ser. **267**, 29 (2023).
[87] K. Cannon *et al.*, arXiv:2010.05082.
[88] S. Privitera, Phys. Rev. D **89**, 024003 (2014).
[89] I. W. Harry, Phys. Rev. D **80**, 104014 (2009).
[90] A. Buonanno, B. Iyer, E. Ochsner, Y. Pan, and B. Sathyaprakash, Phys. Rev. D **80**, 084043 (2009).
[91] J. S. Speagle, Mon. Not. R. Astron. Soc. **493**, 3132 (2020).
[92] S. Biscoveanu, S. Vitale, and C.-J. Haster, Astrophys. J. Lett. **884**, L32 (2019).
[93] LIGO and Virgo Calibration Uncertainty (O1, O2 and O3), https://dcc.ligo.org/T2100313/public (2021).
[94] A. Bohé *et al.*, Phys. Rev. D **95**, 044028 (2017).
[95] B. D. Lackey, M. Pürrer, A. Taracchini, and S. Marsat, Phys. Rev. D **100**, 024002 (2019).