

Efficient reduced order quadrature construction algorithms for fast gravitational wave inference

Gonzalo Morrás^{✉,*}, Jose Francisco Nuño Siles^{✉,†} and Juan García-Bellido[‡]
*Instituto de Física Teórica UAM/CSIC, Universidad Autónoma de Madrid,
 Cantoblanco 28049 Madrid, Spain*

 (Received 1 August 2023; accepted 21 November 2023; published 14 December 2023)

Reduced order quadrature (ROQ) methods can greatly reduce the computational cost of gravitational wave (GW) likelihood evaluations and therefore greatly speed up parameter estimation analyses, which is a vital part to maximize the science output of advanced GW detectors. In this paper, we do an in-depth study of ROQ techniques applied to GW data analysis and present novel algorithms to enhance different aspects of the ROQ bases' construction. We improve upon previous ROQ construction algorithms, allowing for more efficient bases in regions of parameter space that were previously challenging. In particular, we use singular value decomposition methods to characterize the waveform space and choose a reduced order basis close to optimal and also propose improved methods for empirical interpolation node selection, greatly reducing the error added by the empirical interpolation model. To demonstrate the effectiveness of our algorithms, we construct multiple ROQ bases ranging in duration from 4 to 256 s for compact binary coalescence waveforms, including precession and higher order modes. We validate these bases by performing likelihood error tests and percent-percent tests and explore the speedup they induce both theoretically and empirically with positive results. Furthermore, we conduct end-to-end parameter estimation analyses on several confirmed GW events, showing the validity of our approach in real GW data.

DOI: [10.1103/PhysRevD.108.123025](https://doi.org/10.1103/PhysRevD.108.123025)

I. INTRODUCTION

Gravitational wave (GW) astronomy has been made possible in recent years by ground-based observatories like LIGO [1], Virgo [2], and KAGRA [3], revolutionizing our understanding of the Universe by enabling the direct detection of GW signals emitted during extreme cosmic phenomena such as the mergers of binary black holes, binary neutron stars, and neutron star–black hole binaries. With the continuous improvement in sensitivity of current detectors [4] and the advent of next-generation detectors, including projects like the Einstein Telescope [5], Cosmic Explorer [6], and LISA [7–9], we anticipate a dramatic increase in the number of GW candidates detected. For maximum science outputs, a parameter estimation (PE) for each candidate will have to be performed. With standard PE methods [10], this can be prohibitively computationally expensive, especially as we reduce the frequency from which we can detect gravitational waves and the duration of the signals becomes much longer [11].

To fully exploit the enhanced sensitivity of these advanced detectors, it is essential to use accurate waveform models that incorporate important physical effects such as precession or

higher order modes [12]. However, the computational challenge of calculating the likelihood of such signals poses a significant bottleneck in the analysis pipeline. Traditional likelihood calculations can be computationally intensive, particularly for long-duration waveforms. Several methods have been explored in the literature to reduce this computational burden, such as multibanding [13], heterodyned likelihood [14,15], likelihood-free approaches [16,17], reduced order quadrature (ROQ) methods [18–23], and others [24–26].

In this work, we will focus on the ROQ method, which is one of the most promising approaches to fast GW likelihood evaluations, due to its ability to achieve very large speedups while maintaining high accuracy and being able to accommodate the effects of precession and higher order modes [20,21]. ROQ methods exploit the fact that, for a given parameter range, the corresponding GW waveforms span only a small subspace of the vector space of all possible signals. By constructing reduced bases that capture the essential information of the templates, ROQ techniques provide an efficient representation that enables fast likelihood evaluations. The ROQ has a startup cost associated with the off-line basis building stage, which needs to be performed in advance only once per waveform model and parameter space. However, since for typical PE analyses we have to compute more waveforms than what is needed to construct the ROQ and a basis can be used to perform multiple PEs, this startup cost quickly pays off.

*gonzalo.morras@uam.es

†jose.nunno@uam.es

‡juan.garciabellido@uam.es

This paper presents several algorithms for ROQ construction, which offer some key advantages over existing methods. They are specifically designed to tackle the challenges of speed in the basis construction and accuracy in GW likelihood evaluation while maximizing the ROQ speedup. As we will see, these algorithms have the ability to handle complex waveform models in parameter ranges that were intractable with existing procedures.

The paper is organized as follows. In Sec. II, we introduce the basic theoretical framework, including a discussion on GW inference as well as on the basics of ROQ. In Sec. III, we describe the ROQ algorithms we introduce in depth, going through the construction of the reduced order basis, the choice of empirical interpolation model, and how to construct a ROQ with a set tolerance for a given parameter space. In Sec. IV, we present several bases created for two phenomenological waveform models, IMRPhenomPv2 [27] and IMRPhenomXPHM [28], and test their speed and accuracy. We further test the ROQ by performing parameter estimation analyses on three confirmed GW events. In Sec. V, we finally conclude. We relegate some of the more convoluted numerical methods used by our algorithms to the Appendixes.

The methods introduced in this paper have been implemented in a PYTHON code called `EigROQ` [29].

II. THEORETICAL FRAMEWORK

In this section we will briefly describe the basic theoretical framework to contextualize the rest of the paper. In Sec. II A, we give a very brief overview on the basics of GW parameter estimation, and in Sec. II B we summarize the basics of the ROQ rule. For more details, we refer the reader to Refs. [10,20].

A. A primer on gravitational wave inference

GW inference refers to the modern scientific discipline taking care, among other things, of computing the posterior probability distribution of the GW model parameters $\vec{\theta}$ that best fit the data, using the Bayes theorem,

$$p(\vec{\theta}|d) = \frac{\mathcal{L}(d|\vec{\theta})\pi(\vec{\theta})}{\mathcal{Z}}. \quad (1)$$

In this equation, there are several objects that enter the calculation. The first, $\pi(\vec{\theta})$ refers to the prior employed, from the nature of the event, which throughout this paper will always be a compact binary coalescence (CBC) to the distributions describing the parameters of the binary. Next, the likelihood function $\mathcal{L}(d|\vec{\theta})$ of the data given the parameters $\vec{\theta}$ and the evidence \mathcal{Z} represent the probability of the data given the model.

The likelihood is the most computationally expensive part of estimating the posterior. Given a CBC signal without eccentricity, there are 15 different parameters to

fit that enter the likelihood computation. The typical gravitational wave astronomy likelihood is based on the hypothesis that only Gaussian noise is present in the detector and deviations from it are the result of a GW signal. In such case, the likelihood can, up to a normalization constant, be expressed as [30]

$$\begin{aligned} \log \mathcal{L}(d|\vec{\theta}) &= -\frac{1}{2}(d - h(\vec{\theta}), d - h(\vec{\theta})) \\ &= -\frac{1}{2}(d, d) + (d, h(\vec{\theta})) - \frac{1}{2}(h(\vec{\theta}), h(\vec{\theta})), \end{aligned} \quad (2)$$

where $h(\vec{\theta})$ represents, in this specific case, the CBC waveform with parameters $\vec{\theta}$ used to fit the data d . The overlap integral (\cdot, \cdot) is defined as

$$(d, h(\vec{\theta})) = 4\Delta f \mathcal{R} \sum_{j=1}^L \frac{\tilde{d}^*(f_j)\tilde{h}(f_j;\vec{\theta})}{S(f_j)}, \quad (3)$$

where $S_n(f)$ is the detector's noise power spectral density (PSD) and $\tilde{a}(f)$ denotes the Fourier transform of $a(t)$. Since the data of GW detectors are discretely sampled, we will have discrete Fourier transforms having a frequency spacing $\Delta f = 1/T$, with T being the observation time. For a frequency window $(f_{\text{high}} - f_{\text{low}})$ there will be $L = \text{int}[(f_{\text{high}} - f_{\text{low}})T]$ terms in the sum of Eq. (3).¹ Repeatedly computing the overlap integrals in Eq. (2) is the bottleneck in gravitational waves inference and the main part we aim to speed up in this paper.

B. Basics of reduced order quadratures for gravitational wave inference

The parameters $\vec{\theta}$ of the GW signal $h(\vec{\theta})$ we are fitting to the data [Eq. (2)] can be split on intrinsic and extrinsic parameters. The extrinsic parameters are common to all transient GW sources and they are the sky location, usually measured with right ascension α and declination δ , the polarization ψ , luminosity distance d_L , and a reference time of arrival of the signal t_c .² The intrinsic parameters are related to the source of the GW and are generically referred to as $\vec{\lambda}$. For a quasicircular CBC they are composed of the two component masses m_1 and m_2 , three components per black hole spin vector \vec{s}_i , the inclination angle ι , and the coalescence phase ϕ_c . For CBCs with at least one neutron star (NS), $\vec{\lambda}$ can also contain a tidal deformability parameter Λ per NS in the binary [31], as well as any other matter effect information included in the model. If we break the assumption of quasicircular orbits, the eccentricity e would

¹Here $\text{int}[x]$ refers to taking the integer part of x .

²We use t_c because, for the CBC case, the reference time of arrival for the signal is usually given by the coalescence time at the geocenter.

also have to be taken into account in the intrinsic parameters $\vec{\lambda}$ [32].

We assume that the signal $h(t, \vec{\theta})$ is short enough to ignore the dependence of the detector antenna patterns $F_{+, \times}$ with time and the time-varying Doppler shift due to motion of the detector with respect to the Solar System barycenter [33]. In practice, the signal will have to last less than a few hours to be able to ignore the effects of Earth's rotation. Then, in the frequency domain, the GW signal can be written as

$$\begin{aligned} \tilde{h}(f, \vec{\theta}) &= e^{-i2\pi f t_c} \frac{1}{d_L} (F_+(\alpha, \delta, \psi) \tilde{h}_+(f, \vec{\lambda}) \\ &\quad + F_\times(\alpha, \delta, \psi) \tilde{h}_\times(f, \vec{\lambda})) \\ &\equiv e^{-i2\pi f t_c} \tilde{h}(f, \vec{\Lambda}). \end{aligned} \quad (4)$$

The main idea of the ROQ is to represent the GW waveform model $\tilde{h}(f_i; \vec{\theta})$ and its modulus squared $|\tilde{h}(f_i; \vec{\theta})|^2$ in terms of an empirical interpolant each, which is described in more detail in Sec. III. For now, we assume that they can be approximated to arbitrary precision as

$$\tilde{h}(f_i; \vec{\Lambda}) \approx \sum_{j=1}^{N_L} B_j(f_i) \tilde{h}(F_j; \vec{\Lambda}), \quad (5a)$$

$$|\tilde{h}(f_i; \vec{\Lambda})|^2 \approx \sum_{k=1}^{N_Q} C_k(f_i) |\tilde{h}(\mathcal{F}_k; \vec{\Lambda})|^2, \quad (5b)$$

where the main focus of this paper is to find the optimal values of the interpolation nodes $\{F_j\}_{j=1}^{N_L}$ and $\{\mathcal{F}_k\}_{k=1}^{N_Q}$ and of the ‘‘bases’’ $B_j(f_i)$ and $C_k(f_i)$ such that we minimize the required number of elements ($N_L + N_Q$) entering Eq. (5) while respecting a given specified precision.

If we input Eq. (4) into Eq. (2) and use the approximation for the GW waveform $\tilde{h}(f_i; \vec{\theta})$ and its modulus squared $|\tilde{h}(f_i; \vec{\theta})|^2$ of Eq. (5), we can represent the likelihood as

$$\begin{aligned} \log \mathcal{L}(d|\vec{\theta}) &\approx -\frac{1}{2} (d, d) + (d, h(\vec{\theta}))_{\text{ROQ}} \\ &\quad - \frac{1}{2} (h(\vec{\theta}), h(\vec{\theta}))_{\text{ROQ}}, \end{aligned} \quad (6)$$

where the term $-\frac{1}{2}(d, d) \equiv \log \mathcal{L}_{\text{noise}}$ is a constant that depends only on the data and cancels with the evidence \mathcal{Z} when we compute the posterior probability distribution using the Bayes theorem [Eq. (1)]. In Eq. (6) we have also implicitly defined the quantities

$$(d, h(\vec{\theta}))_{\text{ROQ}} \equiv \mathcal{R} \sum_{j=1}^{N_L} w_j(t_c) \tilde{h}(F_j; \vec{\Lambda}), \quad (7a)$$

$$(h(\vec{\theta}), h(\vec{\theta}))_{\text{ROQ}} \equiv \sum_{k=1}^{N_Q} \psi_k |\tilde{h}(\mathcal{F}_k; \vec{\Lambda})|^2, \quad (7b)$$

which approximates the corresponding overlap integrals appearing in the likelihood calculation of Eq. (2). In Eq. (7) we have introduced the linear and quadratic ROQ weights $w_j(t_c)$ and ψ_k , defined as

$$w_j(t_c) \equiv 4\Delta f \sum_{i=1}^L \frac{\tilde{d}^*(f_i) B_j(f_i)}{S(f_i)} e^{-i2\pi f_i t_c}, \quad (8a)$$

$$\psi_k \equiv 4\Delta f \sum_{i=1}^L \frac{C_k(f_i)}{S(f_i)}. \quad (8b)$$

Before starting PE analysis on an event, the weights have to be computed for the observed data strain $\tilde{d}(f)$ and the corresponding PSD $[S(f)]$. Since the linear weights are smooth functions of time, they are usually evaluated in a discrete set of times N_t and are interpolated for the PE analysis [20]. The spacing between time samples is usually on the order of the expected resolution in t_c , which for CBC signals can be as small as 0.1 ms, and for the typical t_c prior, which is uniform in ± 0.1 s around trigger time, this equates to $N_t \sim O(10^3)$. Therefore, at the beginning of the analysis, we have to perform $N_t N_L + N_Q$ full overlaps, as prescribed in Eq. (8), and the startup cost of the ROQ is $O((N_t N_L + N_Q)L)$.

Once the weights have been initialized, computing the ROQ likelihood will only require $N_L + N_Q$ terms to estimate the overlap integrals [Eq. (7)], compared to the L terms in the full overlap integrals. We can therefore expect a speedup in the likelihood computation of $O(L/(N_L + N_Q))$ when using the ROQ rule. In GW astronomy, typical CBC PE analyses require $O(10^8 - 10^9)$ likelihood evaluations, which dominate the computational cost required to sample the posterior of Eq. (1). In most applications, the startup cost of the ROQ is negligible compared to the sampling time and the ROQ will greatly speed up the whole analysis. The likelihood speedup is further explored in Sec. IV C.

The biggest overhead when using the ROQ rule is in constructing the ROQ basis [Eq. (5)], since to explore typical CBC parameter spaces we need $O(10^6 - 10^7)$ random waveforms. With the methods outlined in this paper, we also aim to reduce the computational time of the basis generation, allowing us to handle complex waveform models in parameter ranges that were intractable with existing procedures. In practice, for the CBC case, we train the ROQ on the h_+ polarization, varying only the

values of the intrinsic parameters $\vec{\lambda}$, defined in Eq. (4). The same ROQ basis is valid for both polarizations since they can be jointly decomposed in spherical harmonics of spin weight -2 , ${}_{-2}Y_{lm}$ as [12]

$$h_+ - ih_\times = \sum_{l=2}^{\infty} \sum_{m=-l}^l {}_{-2}Y_{lm}(t, \phi_c) h_{lm}, \quad (9)$$

where the inclination ι and coalescence phase ϕ_c are also being sampled.

III. EFFICIENT ALGORITHM FOR ROQ COMPUTATION

A. Reduced order basis

We generate N templates from the waveform model we are trying to approximate,

$$\{h_A(x), A = 1, \dots, N\}, \quad (10)$$

where, in GW astronomy, x can be either frequency f or time t . We can define the matrix of inner products between templates as

$$M_{AB} = \langle h_A, h_B \rangle. \quad (11)$$

In this context, the inner product is usually defined as

$$\langle h_A, h_B \rangle = \int_{f_{\text{low}}}^{f_{\text{high}}} \tilde{h}_A^*(f) \tilde{h}_B(f) df, \quad (12)$$

although we could also use a reference PSD $S_n(f)$ to give different weights at different frequencies to the integrand, as in Eq. (3). Since M_{AB} is a matrix of inner products, it is Hermitian and positive semidefinite and, therefore, can always be diagonalized as

$$M_{AB} = \sum_{C=1}^N E_{AC} \lambda_C E_{BC}^*, \quad (13)$$

where $\lambda_C \geq 0$ are the eigenvalues and E_{AB} is a unitary matrix whose columns are the orthonormal eigenvectors

$$\sum_{C=1}^N E_{CA}^* E_{CB} = \delta_{AB}. \quad (14)$$

In the waveform space, we can then define the eigenvectors with $\lambda_A \neq 0$ as

$$e_A(x) = \frac{1}{\sqrt{\lambda_A}} \sum_{C=1}^N h_C(x) E_{CA}. \quad (15)$$

It can be proven that these are an orthonormal set of vectors under $\langle \cdot, \cdot \rangle$. That is,

$$\begin{aligned} \langle e_A, e_B \rangle &= \left\langle \frac{1}{\sqrt{\lambda_A}} \sum_{C=1}^N h_C(x) E_{CA}, \frac{1}{\sqrt{\lambda_B}} \sum_{D=1}^N h_D(x) E_{DB} \right\rangle \\ &= \frac{1}{\sqrt{\lambda_A \lambda_B}} \sum_{C=1}^N \sum_{D=1}^N E_{CA}^* E_{DB} \underbrace{\langle h_C, h_D \rangle}_{M_{CD}} \\ &= \frac{1}{\sqrt{\lambda_A \lambda_B}} \sum_{C=1}^N E_{CA}^* \underbrace{\sum_{D=1}^N M_{CD} E_{DB}}_{\lambda_B E_{CB}} \end{aligned} \quad (16)$$

$$= \underbrace{\sqrt{\frac{\lambda_B}{\lambda_A}} \sum_{C=1}^N E_{CA}^* E_{CB}}_{\delta_{AB}} = \delta_{AB}. \quad (17)$$

We can also prove that the inner product between one of the waveforms used to compute M_{AB} and a given eigenvector will be given by

$$\begin{aligned} \langle h_A, e_B \rangle &= \left\langle h_A, \frac{1}{\sqrt{\lambda_B}} \sum_{C=1}^N h_C(x) E_{CB} \right\rangle \\ &= \frac{1}{\sqrt{\lambda_B}} \sum_{C=1}^N \underbrace{\langle h_A, h_C \rangle}_{M_{AC}} E_{CB} \\ &= \frac{1}{\sqrt{\lambda_B}} \sum_{C=1}^N \underbrace{M_{AC} E_{CB}}_{\lambda_B E_{AB}} = \sqrt{\lambda_B} E_{AB}. \end{aligned} \quad (18)$$

We can define our reduced order basis (ROB) as a subset of $n < N$ elements of $\{e_A\}_{A=1}^N$, which we will learn how to optimally select later. To represent the waveform h_A in terms of this ROB $\{e_a\}_{a=1}^n$, we project h_A using the orthonormality property of the ROB,

$$h_A^{\text{ROB}}(x) = \sum_{b=1}^n \langle e_b, h_A \rangle e_b(x) = \sum_{b=1}^n \sqrt{\lambda_b} E_{Ab}^* e_b(x). \quad (19)$$

We can compute the representation error of projecting h_A as

$$\begin{aligned} \sigma_{\text{ROB},A} &= \|h_A - h_A^{\text{ROB}}\|^2 = \langle h_A - h_A^{\text{ROB}}, h_A - h_A^{\text{ROB}} \rangle \\ &= \left\langle h_A - \sum_{b=1}^n \langle e_b, h_A \rangle e_b, h_A - \sum_{c=1}^n \langle e_c, h_A \rangle e_c \right\rangle \\ &= \langle h_A, h_A \rangle - \sum_{b=1}^n |\langle e_b, h_A \rangle|^2 \\ &= \langle h_A, h_A \rangle - \sum_{b=1}^n \lambda_b |E_{Ab}|^2. \end{aligned} \quad (20)$$

Ideally, to construct a ROB we would take a very large number of templates $\{h_A\}_{A=1}^N$ that capture most of the variability of the waveform in the parameter space of interest, compute the matrix M_{AB} as in Eq. (11), diagonalize it, and, to construct our ROB, pick the minimum number of eigenvectors $\{e_a\}_{a=1}^n$ such that the ROB error of Eq. (20) is smaller than a specified tolerance. Unfortunately, this cannot be done in practice, since the number of random templates needed to fully span the typical parameter spaces for GW applications is of order $O(10^7)$. Using the fact that M_{AB} is Hermitian, we need $N(N-1)/2$ complex numbers to store the off-diagonal elements and N real numbers for the diagonal elements. Assuming that each real number is stored with n_B bytes, the memory required to store M_{AB} is³

$$\text{Memory}(M_{AB}) = N^2 n_B = 80 \text{ GB} \left(\frac{N}{10^5} \right)^2 \left(\frac{n_B}{8\text{B}} \right). \quad (21)$$

Therefore, in current computers, examining more than a few tens of thousands of waveforms at a time is unfeasible, and we will not be able to analyze the entire parameter space at once. Motivated by this issue, we have developed a multistep approach summarized in Algorithm 1. We construct a first ROB for a set tolerance with random waveforms. Then, we calculate its orthogonal space and obtain the corresponding ROB, which we add to the original ROB. We repeat this process iteratively, reducing the tolerance at every step. The equivalent to the matrix M_{AB} of Eq. (11) for the orthogonal space to the basis $\{e_a\}_{a=1}^n$ is

$$\begin{aligned} M_{AB}^{\text{ROB}} &= \langle h_A - h_A^{\text{ROB}}, h_B - h_B^{\text{ROB}} \rangle \\ &= \left\langle h_A - \sum_{c=1}^n \langle e_c, h_A \rangle e_c, h_B - \sum_{d=1}^n \langle e_d, h_B \rangle e_d \right\rangle \\ &= \langle h_A, h_B \rangle - \sum_{c=1}^n \langle h_A, e_c \rangle \langle e_c, h_B \rangle. \end{aligned} \quad (22)$$

We observe that the same formulas and reasoning of Eqs. (11)–(21) apply to the space orthogonal to the ROB if we make the identification $h_A \rightarrow h_A - h_A^{\text{ROB}}$. To find the minimum number of elements that have to be added to the ROB to reduce the error below the set tolerance σ , we use Algorithm 2, where we iteratively subtract the contribution of the eigenvalue that produces the largest drop in any $\sigma_{\text{ROB},A}$, according to Eq. (20), until $\sigma_{\text{ROB},A} < \sigma$ for all A .

The process of diagonalizing the matrix M_{AB} of Eq. (11) and finding the eigenvalues in the waveform domain using Eq. (15) is equivalent to performing singular value decomposition (SVD) on a set of waveforms $\{h_A\}_{A=1}^N$, which has been previously used in the literature for the reduced order modeling of GW waveforms (see Refs. [34,35]). However,

³1 GB = 10^9 bytes = 8×10^9 bits.

Algorithm 1. Construction of reduced order basis.

-
- 1: **Input:** Maximum number of waveforms selected N , tolerances of each step $[\sigma_0, \dots, \sigma_s]$, maximum number of waveforms computed per step $[N_{\text{lim},1}, \dots, N_{\text{lim},s}]$
 - 2: Generate N waveforms $\{h_A\}_{A=1}^N$
 - 3: Compute the matrix $M_{AB} = \langle h_A, h_B \rangle$
 - 4: Diagonalize M_{AB} to obtain eigenvalues λ_A and eigenvectors E_{AB}
 - 5: Input $\{\sigma_0, \{h_A\}_{A=1}^N, \lambda_A, E_{AB}\}$ in Algorithm 2 to obtain initial ROQ basis $\{e_i\}_{i=1}^{n_0}$
 - 6: **for** $j = 1 \rightarrow s$ **do**
 - 7: **repeat**
 - 8: Generate $N_{\text{lim},j}$ waveforms $\{h_A\}_{A=1}^{N_{\text{lim},j}}$ and compute their ROB error $\sigma_{\text{ROB},A}$
 - 9: Select the N waveforms $\{h_A\}_{A=1}^N$ with largest σ_{ROB}
 - 10: Save the minimum value of σ_{ROB} for the selected waveforms: $\sigma_{\text{ROB},\min}$
 - 11: $M_{AB}^{\text{ROB}} = \langle h_A, h_B \rangle - \sum_{c=1}^{n_{j-1}} \langle h_A, e_c \rangle \langle e_c, h_B \rangle$
 - 12: Diagonalize M_{AB}^{ROB} and obtain eigenvalues λ_A and eigenvectors E_{AB}
 - 13: Input $\{\sigma_j, \{h_A - h_A^{\text{ROB}}\}_{A=1}^N, \lambda_A, E_{AB}\}$ in Algorithm 2 to obtain next ROQ basis elements $\{e_i\}_{i=n_{j-1}+1}^{n_j}$
 - 14: **until** $\sigma_{\text{ROB},\min} < \sigma_s$
 - 15: **end for**
 - 16: **Output:** ROB $\{e_i\}_{i=1}^n$
-

we follow the procedure outlined in this paper since it has a few numerical advantages. Namely, if we have waveforms with a number of sampling points M , storing them will require $2MNn_B$ bytes, which in the usual case that $M \gg N$, will be much larger than the memory needed to store M_{AB} [Eq. (21)] and we will be even more limited in the number of waveforms we can analyze at once. Moreover, if we are studying the ROB of the space orthogonal to $\{e_a\}_{a=1}^n$, our algorithm is equivalent to computing the SVD of the orthogonal part of the waveforms $\{h_A - h_A^{\text{ROB}}\}_{A=1}^N$. Finding this orthogonal part is, in general, a computationally

Algorithm 2. Selection of eigenvectors.

-
- 1: **Input:** Tolerance σ , waveforms $\{h_A\}_{A=1}^N$, eigenvalues λ_A , and eigenvectors E_{AB} of the matrix $M_{AB} = \langle h_A, h_B \rangle$
 - 2: Initialize $\sigma_A: \{\sigma_A = \langle h_A, h_A \rangle\}_{A=1}^N$
 - 3: Compute the maximum contribution of each eigenvector $\{\delta\sigma_{A,\max} = \lambda_A \max_B |E_{BA}|^2\}_{A=1}^N$
 - 4: Find order of $\delta\sigma_{A,\max}: \{B_n\}_{n=1}^N = \text{argsort}(\delta\sigma_{B,\max})$
 - 5: $n = N$
 - 6: **repeat**
 - 7: Compute current error $\{\sigma_A \leftarrow \sigma_A - \lambda_{B_n} |E_{AB_n}|^2\}_{A=1}^N$
 - 8: $n \leftarrow n - 1$
 - 9: **until** $\sigma_A < \sigma \forall A = 1, \dots, N$
 - 10: **Output:** Eigenvectors in waveform domain $\left\{ e_k(x) = \frac{1}{\sqrt{\lambda_{B_k}}} \sum_{A=1}^N h_A(x) E_{AB_k} \right\}_{k=n}^N$
-

expensive process that can be avoided if M_{AB}^{ROB} is obtained using Eq. (22). Since we are going to select $n_{\text{new}} \ll N$ eigenvectors of M_{AB}^{ROB} , we can just compute the orthogonal projection of their corresponding eigenvectors in the waveform domain at the end of the algorithm.

B. Empirical interpolation model

Writing a given template in the form of Eq. (19) will not save computational cost, since one needs the full waveform $h_A(x)$ to compute the inner product $\langle h_A, e_B \rangle$. To avoid this, we approximate the inner products $\langle h(\vec{\lambda}), e_i \rangle$ by some coefficients $c_i(\vec{\lambda})$ that will, in general, be functions of the parameters of the waveform $\vec{\lambda}$ (e.g., for a CBC this would be masses, spins, inclination, and coalescence phase). The approximate waveform can then be written as

$$I_n[h](x, \vec{\lambda}) = \sum_{i=1}^n c_i(\vec{\lambda}) e_i(x). \quad (23)$$

We force the approximation to be exact at some interpolation nodes $\{X_j\}_{j=1}^m$,

$$I_n[h](X_j, \vec{\lambda}) = h(X_j, \vec{\lambda}) = \sum_{i=1}^n c_i(\vec{\lambda}) e_i(X_j). \quad (24)$$

This is what we define as an interpolant. If we identify the matrix

$$A_{ij} = e_j(X_i), \quad (25)$$

and take the number of interpolation nodes m to be equal to the number of basis elements n , then \hat{A} is a square matrix that we construct by choosing the interpolation nodes $\{X_j\}_{j=1}^n$. Assuming that we construct \hat{A} to be invertible, we can solve Eq. (24) for $c_i(\vec{\lambda})$ in the following way:

$$c_i(\vec{\lambda}) = \sum_{j=1}^n (\hat{A}^{-1})_{ij} h(X_j, \vec{\lambda}). \quad (26)$$

We therefore observe that the value of $c_i(\vec{\lambda})$ will just be a linear combination of the values of the waveform at the different interpolation nodes $\{X_j\}_{j=1}^n$. In practice, the functions $h(x)$ and the ROB elements $\{e_i(x)\}_{i=1}^n$ are discretely sampled in a set of points $\{x_i\}_{i=1}^M$, and we can define the matrix

$$\hat{V} \equiv [\vec{e}_1, \dots, \vec{e}_n] \in \mathbb{C}^{M \times n}, \quad (27)$$

where $\vec{e}_A = e_A(\vec{x}) \in \mathbb{C}^M$. From Eq. (25), we observe that the matrix \hat{A} can be written in terms of \hat{V} as

$$\hat{A} = \hat{P}^\dagger \hat{V} \in \mathbb{C}^{n \times n}, \quad (28)$$

where the matrix $\hat{P} \in \mathbb{C}^{M \times n}$ is a projector that selects the rows of \hat{V} corresponding to the interpolation nodes. That is,

$$P_{aj} = \delta_{aj}, \quad (29)$$

with $\{\beta_j\}_{j=1}^n$ as the indices of the interpolation nodes (i.e., $x_{\beta_j} = X_j$). In terms of these matrices, the empirical interpolation model (EIM) can be written as

$$I_n[\vec{h}] = \hat{V}(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger \vec{h}, \quad (30)$$

which is an interpolant because $\hat{P}^\dagger I_n[\vec{h}] = \hat{P}^\dagger \vec{h}$. In terms of the matrix \hat{V} , the ROB representation of \vec{h} is given by

$$\vec{h}^{\text{ROB}} = \hat{V} \hat{V}^\dagger \vec{h}. \quad (31)$$

Note that, even though the basis elements \vec{e}_A are orthonormal, and therefore $\hat{V}^\dagger \hat{V} = \mathbb{1}_{n \times n}$, since the matrices are not square, we have that, in general, $\hat{V} \hat{V}^\dagger \neq \mathbb{1}_{M \times M}$. From Eqs. (30) and (31), we can explicitly see that the EIM acting on a waveform in the ROB space will have no effect. That is,

$$\begin{aligned} I_n[\vec{h}^{\text{ROB}}] &= \hat{V}(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger (\hat{V} \hat{V}^\dagger \vec{h}) = \hat{V}(\hat{P}^\dagger \hat{V})^{-1} (\hat{P}^\dagger \hat{V}) \hat{V}^\dagger \vec{h} \\ &= \hat{V} \hat{V}^\dagger \vec{h} = \vec{h}^{\text{ROB}}. \end{aligned} \quad (32)$$

This can be used to relate the representation error of the EIM with the representation error of the ROB. Computing the modulus of the difference between the exact waveform and its EIM representation, we obtain

$$\begin{aligned} \sigma_{\text{EIM}}(\vec{h}) &= \|\vec{h} - I_n[\vec{h}]\|^2 = \|\mathbb{1} - \hat{V}(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger\| \vec{h}\|^2 \\ &= \|\mathbb{1} - \hat{V}(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger\| (\vec{h} - \vec{h}^{\text{ROB}})\|^2 \\ &\leq \|\mathbb{1} - \hat{V}(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger\|_2^2 \underbrace{\|\vec{h} - \vec{h}^{\text{ROB}}\|^2}_{\sigma_{\text{ROB}}(\vec{h})}, \end{aligned} \quad (33)$$

where $\|\cdot\|_2$ denotes the matrix two-norm, which is given by

$$\begin{aligned} \|\hat{M}\|_2 &= \max_{\vec{x} \neq 0} \frac{\|\hat{M} \vec{x}\|}{\|\vec{x}\|} = \sqrt{\lambda_{\max}(\hat{M}^\dagger \hat{M})} \\ &= \sqrt{\lambda_{\max}(\hat{M} \hat{M}^\dagger)}, \end{aligned} \quad (34)$$

where $\|\vec{x}\|$ is the usual vector norm and $\lambda_{\max}(\hat{M}^\dagger \hat{M})$ denotes the maximum eigenvalue of $\hat{M}^\dagger \hat{M}$. Since $\hat{V}(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger$ is idempotent, that is $(\hat{V}(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger)^2 = \hat{V}(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger$, and it is different from 0 or the identity 1, it follows that [36]

$$\|\mathbb{1} - \hat{V}(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger\|_2 = \|\hat{V}(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger\|_2. \quad (35)$$

Furthermore, since $\hat{V}^\dagger \hat{V} = \mathbb{1}_{n \times n}$ and $\hat{P}^\dagger \hat{P} = \mathbb{1}_{n \times n}$, from the definition in Eq. (34) of the matrix two-norm, we have that

$$\|\hat{V}(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger\|_2 = \|(\hat{P}^\dagger \hat{V})^{-1}\|_2. \quad (36)$$

Substituting in Eq. (33),

$$\sigma_{\text{EIM}}(\vec{h}) \leq \|(\hat{P}^\dagger \hat{V})^{-1}\|_2^2 \sigma_{\text{ROB}}(\vec{h}) = \|\hat{A}^{-1}\|_2^2 \sigma_{\text{ROB}}(\vec{h}). \quad (37)$$

Using the definition of the matrix two-norm of Eq. (34), we have that

$$\begin{aligned}\|\hat{A}^{-1}\|_2^2 &= \lambda_{\max}((\hat{A}^{-1})^\dagger \hat{A}^{-1}) = \lambda_{\max}((\hat{A}^\dagger)^{-1} \hat{A}^{-1}) \\ &= \lambda_{\max}((\hat{A}\hat{A}^\dagger)^{-1}) = \frac{1}{\lambda_{\min}(\hat{A}\hat{A}^\dagger)},\end{aligned}\quad (38)$$

and we can rewrite Eq. (37) as

$$\sigma_{\text{EIM}}(\vec{h}) \leq \frac{\sigma_{\text{ROB}}(\vec{h})}{\lambda_{\min}(\hat{A}\hat{A}^\dagger)}.\quad (39)$$

Therefore, given a maximum error of the ROB, the error of the EIM model is bounded from above by Eq. (39). To make this bound as stringent as possible, we could maximize the smallest eigenvalue of $\hat{A}\hat{A}^\dagger$. Using the definition of \hat{A} from Eq. (28), we can write

$$(\hat{A}\hat{A}^\dagger)_{ij} = \sum_{k=1}^n e_k(X_i) e_k^*(X_j) = \langle \vec{v}_j, \vec{v}_i \rangle, \quad (40)$$

where we have defined the vectors $\{\vec{v}_i\}_k = e_k(X_i) |k = 1, \dots, n\}_{i=1}^M$ as the rows of \hat{V} corresponding to the interpolation nodes X_i . We then observe that $\hat{A}\hat{A}^\dagger$ is the same as the scalar product between the corresponding selected rows of \hat{V} .

If the vectors \vec{v}_i were orthonormal, we would obtain that $(\hat{A}\hat{A}^\dagger)_{ij} = \delta_{ij}$, and therefore $\lambda_{\min}(\hat{A}\hat{A}^\dagger) = 1$ and the EIM would not introduce additional error over the ROB. Selecting n orthonormal rows of \hat{V} is, in general, not possible; however, we can try to minimize the EIM error by picking rows that are as close to orthogonal as possible using Algorithm 3.

We observe that Algorithm 3 is equivalent to picking the EIM nodes that maximize the determinant of $\hat{A}\hat{A}^\dagger$, since

$$\begin{aligned}\det(\hat{A}\hat{A}^\dagger) &= \det(\hat{A}) \det(\hat{A}^\dagger) = |\det(\hat{A})|^2 \\ &= \prod_{j=1}^n |\langle \vec{w}_j, \vec{v}_{\beta_j} \rangle|^2.\end{aligned}\quad (41)$$

Algorithm 3. Selection of interpolation nodes.

-
- 1: **Input:** Evaluated basis $\{\vec{e}_i\}_{i=1}^n$
 - 2: Define row vectors: $\{\vec{v}_\alpha = \{e_i(x_\alpha)\}_{i=1}^n\}_{\alpha=1}^M$
 - 3: Initialize orthonormal base (OB) of columns: $\text{OB} = \{\vec{w}_i\}_{i=1}^0$
 - 4: Initialize the norm of the orthogonal part of \vec{v}_α to OB: $\{N_\alpha = |\vec{v}_\alpha|^2\}_{\alpha=1}^M$
 - 5: **for** $j = 1 \rightarrow n$ **do**
 - 6: Choose vector with largest N_α : $\beta_j = \text{argmax}(N_\alpha)$
 - 7: Append \vec{v}_{β_j} to OB using the Gram-Schmidt process
 - 8: Update N_α : $\{N_\alpha \leftarrow N_\alpha - |\langle \vec{w}_j, \vec{v}_\alpha \rangle|^2\}_{\alpha=1}^M$
 - 9: **end for**
 - 10: **Output:** EIM interpolation nodes $\{\beta_i\}_{i=1}^n$
-

Algorithm 3 does not directly maximize the minimum eigenvalue of $\hat{A}\hat{A}^\dagger$. However, based on the expression for the determinant of $\hat{A}\hat{A}^\dagger$,

$$\det(\hat{A}\hat{A}^\dagger) = \prod_{j=1}^n \lambda_j, \quad (42)$$

to maximize it, the values of the individual eigenvalues have to be large, and thus, the output of the algorithm is near to the minimum of $\|\hat{A}^{-1}\|_2^2$. When compared to the greedy algorithm typically used in the literature (e.g., Refs. [18–21]) to compute the interpolation nodes, we observe a superior performance of Algorithm 3, as we will later discuss in relation to Fig. 1.

If we wanted to create an EIM with a tolerance smaller than σ , from Eq. (39) we could, in principle, just construct a ROB with a tolerance better than $\lambda_{\min}(\hat{A}\hat{A}^\dagger)\sigma$. However, in real settings, we observe that Eq. (39) is a loose upper bound on the EIM error, and we can obtain an EIM with a tolerance better than σ using fewer basis elements.

Instead of bounding $\sigma_{\text{EIM}}(\vec{h})$ using the inequality of Eq. (33), we can refine this expression by doing

$$\begin{aligned}\sigma_{\text{EIM}}(\vec{h}) &= \|\mathbb{1} - \hat{V}(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger\|(\vec{h} - \vec{h}^{\text{ROB}})\|^2 \\ &= \|\vec{h} - \vec{h}^{\text{ROB}}\|^2 + \|\hat{V}(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger(\vec{h} - \vec{h}^{\text{ROB}})\|^2 \\ &= \sigma_{\text{ROB}}(\vec{h}) + \|(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger(\vec{h} - \vec{h}^{\text{ROB}})\|^2,\end{aligned}\quad (43)$$

where we have used that $\hat{V}^\dagger \hat{V} = \mathbb{1}$ and that the EIM projects the waveform onto the ROB, and therefore $\langle \hat{V}(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger(\vec{h} - \vec{h}^{\text{ROB}}), \vec{h} - \vec{h}^{\text{ROB}} \rangle = 0$. From Eq. (43) we have that the EIM error is always larger than or equal to the ROB error. We also observe that for the bound of Eq. (39) to be saturated we need $\hat{P}^\dagger(\vec{h} - \vec{h}^{\text{ROB}})$ to be the eigenvector of $\hat{A}^\dagger \hat{A}$ with the maximum eigenvalue, which is extremely unlikely in general. To explore this, we assume that $\vec{h} - \vec{h}^{\text{ROB}} \equiv \vec{\delta h}$ is a random variable, such that

$$\mathbb{E}[\delta h_\alpha^* \delta h_\beta] = c_\alpha \delta_{\alpha\beta}, \quad (44)$$

where $\mathbb{E}[\cdot]$ denotes the expected value (i.e., the average over random waveform realizations). Using Eq. (44), we compute the expected value of σ_{EIM} as

$$\begin{aligned}\mathbb{E}[\sigma_{\text{EIM}}] &= \sum_{\alpha=1}^M \mathbb{E}[\delta h_\alpha^* \delta h_\alpha] \\ &\quad + \sum_{k=1}^n \sum_{q=1}^n \sum_{l=1}^n (A^{-1})_{lk}^* (A^{-1})_{lq} \mathbb{E}[\delta h_{\beta_k}^* \delta h_{\beta_q}] \\ &= \sum_{\alpha=1}^M c_\alpha + \sum_{k=1}^n \sum_{l=1}^n c_{\beta_k} |(A^{-1})_{lk}|^2 \\ &= \sum_{\alpha=1}^M c_\alpha + \|\hat{A}^{-1}\|_F^2,\end{aligned}\quad (45)$$

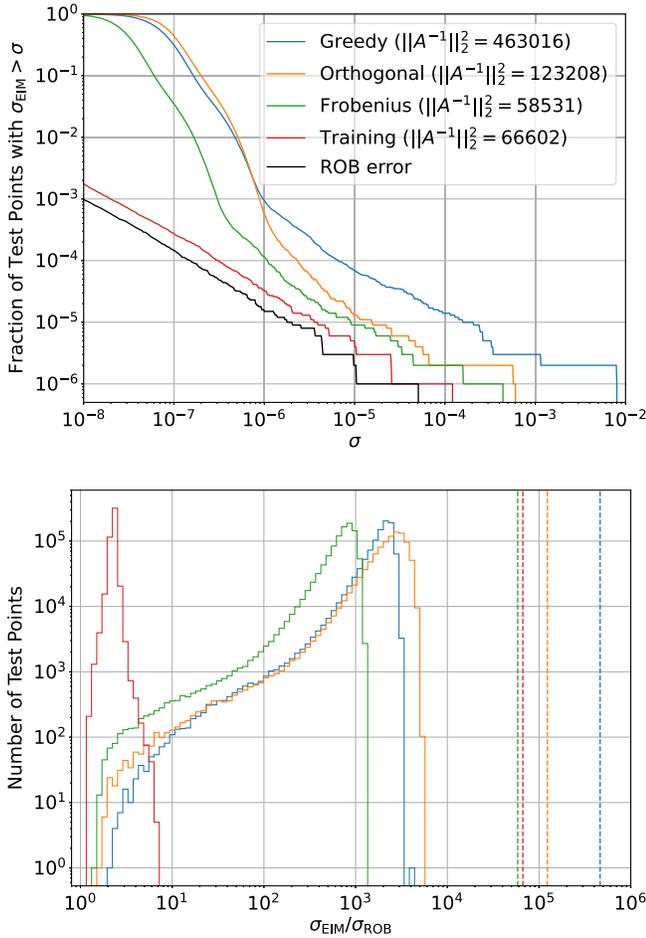


FIG. 1. Comparison of methods to compute the EIM for the 256 sIMRPhenomPv2 ROB of Table I. We test the different EIMs on the same 10^6 samples randomly drawn from the parameter space over which the ROB is generated (see Table I). The “greedy” method is the one outlined in [19], the “orthogonal” method stands for Algorithm 3, the “Frobenius” method corresponds to using Algorithm 4 to minimize $\|\hat{A}^{-1}\|_F$, and the “training” method is the one used to construct the EIM of Table I with Algorithm 5. Upper: fraction of samples with an EIM error larger than a tolerance σ as a function of σ . For comparison purposes, we also show the distribution of the ROB error. Lower: histogram of the ratio between the EIM error and the ROB error for the same methods and test samples as in the upper figure. The vertical dashed lines represent an upper bound, defined by the value of $\|\hat{A}^{-1}\|_2$ for each method.

where \hat{A} is the matrix

$$\tilde{A}_{kl} = \frac{1}{\sqrt{c_{\beta_k}}} A_{kl} = \frac{1}{\sqrt{c_{\beta_k}}} e_l(x_{\beta_k}), \quad (46)$$

such that $(\tilde{A}^{-1})_{lk} = \sqrt{c_{\beta_k}} (A^{-1})_{lk}$ and $\|\cdot\|_F$ is the Frobenius norm, defined as

$$\|\hat{M}\|_F = \sqrt{\sum_{k=1}^n \sum_{l=1}^n |M_{kl}|^2} = \sqrt{\text{Tr}\{\hat{M}^\dagger \hat{M}\}}. \quad (47)$$

Therefore, to optimize the EIM such that the expected value of σ_{EIM} is minimum, we want to minimize the value of the Frobenius norm of \hat{A}^{-1} . Using the properties of the trace, we can rewrite it as

$$\|\hat{A}^{-1}\|_F = \sqrt{\sum_{k=1}^n \frac{1}{\lambda_k(\hat{A}^\dagger \hat{A})}}. \quad (48)$$

To minimize the Frobenius norm of \hat{A} we can start from the EIM given by Algorithm 3 and allow the interpolation nodes to “walk” in the direction of diminishing $\|\hat{A}^{-1}\|_F$, as outlined in Algorithm 4.

The time complexity of Algorithm 4 is $O(N_{\text{rounds}} n N_F)$, where N_F denotes the number of operations required to compute $F(\vec{\beta})$. Given that our target function is $F(\vec{\beta}) = \|\hat{A}^{-1}\|_F$, one could naively expect that, based on the size $n \times n$ of the matrix \hat{A} , directly inverting it would take $O(n^3)$ operations, and therefore the time complexity of Algorithm 4 would be $O(N_{\text{rounds}} n^4)$. This can be computationally very expensive even if $n \ll M$. However, updating the value of $\|\hat{A}^{-1}\|_F$ when only one row of the matrix changes can be done in $O(n^2)$ by following the procedure in the Appendix, and we can implement Algorithm 4 with target function $F(\vec{\beta}) = \|\hat{A}^{-1}\|_F$ in a way that takes $O(N_{\text{rounds}} n^3)$ operations.

Even though Algorithm 4 is considerably better than the greedy algorithms used in the literature, as we will later

Algorithm 4. Selection of interpolation nodes to minimize target function of the EIM $F(\cdot)$.

-
- 1: **Input:** Maximum number of rounds N_{rounds} , initial interpolation nodes $\vec{\beta}$, function to be minimized $F(\vec{\beta})$.
 - 2: **for** $j = 1 \rightarrow N_{\text{rounds}}$ **do**
 - 3: **for** $k = 1 \rightarrow n$ **do**
 - 4: **for** $\delta\beta$ in $[-1, 1]$ **do**
 - 5: Copy interpolation nodes: $\vec{\beta}' = \vec{\beta}$
 - 6: **repeat**
 - 7: Test new EIM: $\beta'_k \leftarrow \beta'_k + \delta\beta$
 - 8: **if** $F(\vec{\beta}') \leq F(\vec{\beta})$ **then**
 - 9: Update reference EIM: $\vec{\beta} \leftarrow \vec{\beta}'$
 - 10: **end if**
 - 11: **until** $F(\vec{\beta}') > F(\vec{\beta})$
 - 12: **end for**
 - 13: **end for**
 - 14: **if** $\{\beta_i\}_{i=1}^n$ did not change this iteration **then**
 - 15: **break for loop**
 - 16: **end if**
 - 17: **end for**
 - 18: **Output:** EIM interpolation nodes $\{\beta_i\}_{i=1}^n$
-

discuss in relation to Fig. 1, it can still be improved by training the EIM directly on the waveform data. For this purpose, we assume that we have an initial ROB $\{\vec{e}_i\}_{i=1}^n$ with a corresponding EIM that can be computed with, e.g., Algorithm 4. We want to update this EIM to better fit a training set of waveforms $\{h_A\}_{A=1}^N$. We first generate a ROB for the part of the training set orthogonal to the initial ROB ($\vec{h} - \vec{h}^{\text{ROB}}$), which can be done by diagonalizing the matrix of Eq. (22). Analogously to Eq. (19), we can write

$$\vec{h}_A - \vec{h}_A^{\text{ROB}} = \sum_{B=1}^N \sqrt{\lambda_B} E_{AB}^* \vec{u}_B, \quad (49)$$

where λ_b and E_{AB} are the eigenvalues and eigenvectors of the matrix M_{AB}^{ROB} defined in Eq. (22) and \vec{u}_B represent the eigenvectors in the waveform domain. Substituting Eq. (49) in the expression for σ_{EIM} derived in Eq. (43), we obtain

$$\begin{aligned} \sigma_{\text{EIM},A} &= \|\vec{h}_A - I_n[\vec{h}_A]\|^2 \\ &= \left\| \sum_{B=1}^N \sqrt{\lambda_B} E_{AB}^* \vec{u}_B \right\|^2 \\ &\quad + \left\| \sum_{B=1}^N \sqrt{\lambda_B} E_{AB}^* (\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger \vec{u}_B \right\|^2 \\ &= \sum_{B=1}^N \lambda_B |E_{AB}|^2 \\ &\quad + \sum_{B=1}^N \sum_{C=1}^N \sqrt{\lambda_B \lambda_C} E_{AB}^* E_{AC} \langle \vec{w}_C, \vec{w}_B \rangle, \end{aligned} \quad (50)$$

where we have defined

$$\vec{w}_B = (\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger \vec{u}_B. \quad (51)$$

From Eq. (50), we can compute the sum of all the EIM errors of the waveforms in the training set. That is,

$$\begin{aligned} \sigma_{\text{EIM}}^{\text{tot}} &= \sum_{A=1}^N \sigma_{\text{EIM},A} = \sum_{B=1}^N \lambda_B (1 + \langle \vec{w}_B, \vec{w}_B \rangle) \\ &= \sum_{B=1}^N \lambda_B (1 + \|(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger \vec{u}_B\|^2) \\ &\approx \sum_{B=1}^{n_\lambda} \lambda_B (1 + \|(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger \vec{u}_B\|^2), \end{aligned} \quad (52)$$

where we have used that E_{AB} is unitary and that the matrix M_{AB}^{ROB} will usually have a small number of large eigenvalues, with the rest of the eigenvalues close to 0. Therefore, we can truncate the sum to be made only over the largest n_λ eigenvalues and obtain a very good approximation of $\sigma_{\text{EIM}}^{\text{tot}}$.

To minimize the value of $\sigma_{\text{EIM}}^{\text{tot}}$, we follow Algorithm 5, in which we start with an EIM and perform walks around the initial solution in the direction of diminishing $\sigma_{\text{EIM}}^{\text{tot}}$. For the initial solution, we will use the EIM generated by Algorithm 4 with target function $F(\vec{\beta}) = \|\hat{A}^{-1}\|_F$. Since we want to fit $\{h_A\}_{A=1}^N$, following Eq. (44), the weights c_α of Eq. (46) are

$$\begin{aligned} c_\alpha &= \mathbb{E}[\delta h_\alpha^* \delta h_\alpha] = \frac{1}{N} \sum_{A=1}^N |h_{A,\alpha} - h_{A,\alpha}^{\text{ROB}}|^2 \\ &= \frac{1}{N} \sum_{A=1}^N \sum_{B=1}^N \sum_{C=1}^N \sqrt{\lambda_B \lambda_C} E_{AB}^* E_{AC} u_{C,\alpha}^* u_{B,\alpha} \\ &= \frac{1}{N} \sum_{B=1}^N \lambda_B |u_{B,\alpha}|^2 \approx \frac{1}{N} \sum_{B=1}^{n_\lambda} \lambda_B |u_{B,\alpha}|^2, \end{aligned} \quad (53)$$

where we have once again used that E_{AB} is unitary and that the sum can be approximated by taking only the largest n_λ eigenvalues. In Algorithm 4, using $\sigma_{\text{EIM}}^{\text{tot}}$ as target function, the value of $\sigma_{\text{EIM}}^{\text{tot}}$ can be efficiently updated with $O(nn_\lambda)$ operations, as described in the Appendix. Therefore, using Algorithm 4 to walk around an initial solution minimizing $\sigma_{\text{EIM}}^{\text{tot}}$ will require $O(N_{\text{rounds}} n^2 n_\lambda)$ operations.

In Fig. 1, we show for the 256 s IMRPhenomPv2 ROB listed in Table I a comparison between Algorithms 3–5 proposed in this paper, the usual greedy algorithm used in the literature, and the lower bound imposed by the ROB error. We show only the analysis for the 256 s IMRPhenomPv2 basis of Table I, but we find similar results for all the other cases in Tables I and II. In the upper panel of Fig. 1 we show the fraction of points with an EIM error larger than a tolerance σ as a function of σ . Comparing the methods, we observe that the training one (Algorithm 5) outperforms the others, which is expected since it has been

Algorithm 5. Selection of interpolation nodes trained on a set of waveforms $\{h_A\}_{A=1}^N$.

- 1: **Input:** Evaluated basis $\{\vec{e}_i\}_{i=1}^n$, maximum number of rounds N_{rounds} , n_λ eigenvalues λ_B , and eigenvectors in waveform domain \vec{u}_B of the matrix $M_{AB}^{\text{ROB}} = \langle \vec{h}_A - \vec{h}_A^{\text{ROB}}, \vec{h}_B - \vec{h}_B^{\text{ROB}} \rangle$.
- 2: Compute weights: $c_\alpha = \sum_{B=1}^{n_\lambda} \lambda_B |u_{B,\alpha}|^2$
- 3: Compute weighted basis: $\{\{w_i(x_\alpha)\}_{\alpha=1}^n\}_{i=1}^M = \left\{ \left\{ \frac{1}{\sqrt{c_\alpha}} e_i(x_\alpha) \right\}_{\alpha=1}^n \right\}_{i=1}^M$
- 4: Get initial EIM $\vec{\beta}$ inputting $\{\vec{w}_i\}_{i=1}^M$ in Algorithm 3
- 5: Update $\vec{\beta}$ using Algorithm 4 with maximum rounds N_{rounds} and target function $F(\vec{\beta}) = \|\hat{A}^{-1}\|_F$, where $\tilde{A}_{ij} = w_j(x_{\beta_i})$
- 6: Update $\vec{\beta}$ again with Algorithm 4 with maximum rounds N_{rounds} and target function $F(\vec{\beta}) = \sigma_{\text{EIM}}^{\text{tot}}$, where $\sigma_{\text{EIM}}^{\text{tot}} = \sum_{B=1}^{n_\lambda} \lambda_B (1 + \sum_{i=1}^n |\sum_{j=1}^M (\hat{A}^{-1})_{ij} u_{B,\beta_j}|^2)$ and $A_{ij} = e_j(x_{\beta_i})$
- 7: **Output:** EIM interpolation nodes $\{\beta_i\}_{i=1}^n$

TABLE I. Summary of the reduced bases constructed with EigROQ for the IMRPhenomPv2 waveform model. We limit the mass ratio $1 \leq q \leq 8$, the magnitudes of the two spins $-0.8 \leq \chi_i \leq 0.8$ for $i \in [1, 2]$, and the full range for the spin angles $(0, 0) \leq (\theta_j, \alpha_0) \leq (\pi, 2\pi)$. For the first base ($\Delta f = 0.25$ Hz) we extend the coverage in spins to $-0.88 \leq \chi_i \leq 0.88$. For the creation of all the bases, we run EigROQ with the same configuration. In Algorithm 1, we set the maximum number of waveforms selected $N = 20000$, tolerances of each step $\sigma_i = [10^{-1}, 10^{-3}, 10^{-5}]$, and maximum number of waveforms computed per step $N_{\text{lim},i} = [10^6, 3.16 \times 10^6]$. In Algorithm 6, we set $N = 10^7$, $\sigma = 10^{-5}$, $N_{\text{lim}} = 10^7$, and the maximum number of eigenvectors used $n_\lambda = 5000$, except for the 256 s basis, where we set $n_\lambda = 4000$ due to memory limitations. The bases are tested on 10^7 randomly generated waveforms in the same parameter space on which the training was done. The ‘‘Theoretical’’ speedup has been computed with Eq. (56), while the ‘‘Empirical’’ speedup is the median and 90% credible interval of the corresponding points in the upper panel of Fig. 5.

Frequency range (Hz)			$M_c(M_\odot)$		Basis size		Test set $\sigma_{\text{EI,max}}$		Test set $\sigma_{\text{EI}} > 10^{-5}$		Likelihood speedup	
Min.	Max.	Δf (Hz)	Min.	Max.	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic	Theoretical	Empirical
20	1024	1/4	12.3	45	242	194	1.00×10^{-3}	1.09×10^{-4}	31	19	9.2	$3.7^{+1.0}_{-0.6}$
20	1024	1/8	7.93	14.76	369	294	4.91×10^{-4}	1.46×10^{-4}	55	31	12.1	$7.1^{+0.7}_{-0.1}$
20	2048	1/16	5.14	9.52	493	389	6.85×10^{-4}	5.72×10^{-4}	110	59	36.8	$22.3^{+0.6}_{-1.4}$
20	2048	1/32	3.35	6.17	631	438	6.88×10^{-4}	5.83×10^{-4}	98	75	60.7	$38.1^{+0.5}_{-0.4}$
20	2048	1/64	2.18	4.02	848	407	1.51×10^{-3}	5.71×10^{-4}	103	71	103.4	$65.7^{+1.6}_{-0.9}$
20	4096	1/128	1.42	2.60	1315	306	6.4×10^{-4}	2.46×10^{-3}	83	50	321.9	$232.3^{+8.0}_{-7.0}$
20	4096	1/256	0.95	1.72	2196	300	1.43×10^{-4}	6.32×10^{-5}	69	28	418.1	$350.7^{+49.8}_{-17.8}$

TABLE II. Summary of the reduced bases constructed with EigROQ for the IMRPhenomXPHM waveform model. We limit the mass ratio $1 \leq q \leq 4$, the magnitudes of the two spins $-0.8 \leq \chi_i \leq 0.8$ for $i \in [1, 2]$, and the full range for the spin angles $(0, 0) \leq (\theta_j, \alpha_0) \leq (\pi, 2\pi)$. For the creation of all the basis, we run EigROQ with the same configuration. In Algorithm 1, we set the maximum number of waveforms selected $N = 20000$, tolerances of each step $\sigma_i = [10^{-2}, 10^{-3}, 10^{-4}]$, and maximum number of waveforms computed per step $N_{\text{lim},i} = [10^6, 3.16 \times 10^6]$. In Algorithm 6, we set $N = 10^7$, $\sigma = 10^{-4}$, $N_{\text{lim}} = 10^7$, and the maximum number of eigenvectors used $n_\lambda = 5000$. The basis are tested on 10^7 randomly generated waveforms in the same parameter space on which the training was done. The theoretical (Th.) speedup has been computed with Eq. (56), while the empirical (Emp.) speedup is the median and 90% credible interval of the corresponding points in the lower panel of Fig. 5. For the empirical speedups, we show the values both without (Emp.) and with (MB) the IMRPhenomXPHM multibanding option enabled [37].

Frequency range (Hz)			$M_c(M_\odot)$		Basis size		Test set $\sigma_{\text{EI,max}}$		Test set $\sigma_{\text{EI}} > 10^{-4}$		Likelihood speedup		
Min.	Max.	Δf (Hz)	Min.	Max.	Linear	Quadratic	Linear	Quadratic	Linear	Quadratic	Theoretical	Empirical	MB
20	1024	1/4	55	110	303	195	3.67×10^{-2}	2.47×10^{-2}	119	86	8.1	$3.2^{+1.2}_{-0.6}$	$1.4^{+0.3}_{-0.3}$
20	1024	1/4	35	66	339	192	6.95×10^{-2}	2.47×10^{-2}	115	64	7.6	$4.5^{+1.7}_{-1.0}$	$1.7^{+0.5}_{-0.3}$
20	1024	1/4	26	42	328	204	9.57×10^{-3}	1.04×10^{-2}	84	21	7.6	$6.1^{+1.8}_{-1.1}$	$2.2^{+0.6}_{-0.5}$
20	1024	1/4	18	33	348	201	1.80×10^{-2}	1.32×10^{-3}	70	19	7.3	$7.8^{+0.4}_{-1.7}$	$2.6^{+0.7}_{-0.6}$
20	1024	1/4	12	20	371	264	1.18×10^{-2}	1.03×10^{-3}	67	16	6.3	$7.5^{+0.3}_{-1.6}$	$3.1^{+0.7}_{-0.6}$
20	1024	1/8	10	15	491	386	4.32×10^{-3}	4.39×10^{-4}	50	6	9.2	$11.1^{+0.3}_{-0.4}$	$4.3^{+1.2}_{-1.0}$
20	1024	1/8	8.6	11.8	505	435	9.33×10^{-3}	1.96×10^{-4}	48	3	8.5	$10.5^{+0.3}_{-0.9}$	$4.8^{+0.8}_{-1.0}$
20	2048	1/16	5.1	9.6	868	942	2.95×10^{-3}	2.38×10^{-3}	56	11	17.9	$24.6^{+2.3}_{-4.6}$	$4.8^{+1.2}_{-0.8}$
20	2048	1/32	3.35	6.17	1539	1826	9.62×10^{-4}	2.53×10^{-4}	46	1	19.3	$27.6^{+1.1}_{-0.8}$	$4.6^{+1.8}_{-0.9}$
20	2048	1/64	2.18	4.02	2924	3636	6.37×10^{-4}	2.68×10^{-4}	19	7	19.8	$28.6^{+0.7}_{-0.5}$	$4.2^{+1.7}_{-0.7}$

trained on the waveform data to reduce the EIM error. The worst performer is the greedy method, since it induces the largest EIM error in all cases tested. We also observe that the Frobenius method, which uses Algorithm 4 to minimize $\|\hat{A}^{-1}\|_F$, induces the smallest EIM error among the algorithms that do not train on waveforms, which could make it more robust against overfitting.

In the lower panel of Fig. 1, we show the ratio between the EIM and the ROB error for the same methods and test samples

as in the upper panel. We observe that this ratio is in the range $1 \leq \sigma_{\text{EIM}}/\sigma_{\text{ROB}} \leq \|\hat{A}^{-1}\|_2^2$, as was derived in Eqs. (37) and (43). In general, we observe that the EIM errors obtained with the different methods are always considerably below the upper limit imposed by Eq. (37) ($\sigma_{\text{EIM}}/\sigma_{\text{ROB}} \ll \|\hat{A}^{-1}\|_2^2$). This is expected, since to saturate this upper bound we need $\hat{P}^\dagger(\vec{h} - \vec{h}^{\text{ROB}})$ to be the eigenvector of $\hat{A}^\dagger \hat{A}$ with the maximum eigenvalue, which is hard to get in practice. We also

observe that the training method is almost optimal since most samples are close to the lower bound of $\sigma_{\text{EIM}}/\sigma_{\text{ROB}} \geq 1$. In contrast, most of the samples for the methods that do not involve training on waveform data concentrate at values of $\sigma_{\text{EIM}}/\sigma_{\text{ROB}} \gtrsim 10^3$. This is probably because, when we train on the waveform data, we are selecting an EIM that avoids coincidences between $\hat{P}^\dagger(\vec{h} - \vec{h}^{\text{ROB}})$ and eigenvectors of $\hat{A}^\dagger \hat{A}$ with large eigenvalues.

C. Construction of the ROQ

In this section, we describe how we use the methods of Secs. III A and III B to create, in an efficient way, an EIM that fits a waveform model over a parameter space with a tolerance better than σ .

We obtain an initial ROB $\{\vec{e}_i\}_{i=1}^n$ using Algorithm 1 and construct its corresponding EIM with Algorithm 5, where the set of training waveforms is the $\{h_A\}_{A=1}^N$, selected in the last step of Algorithm 1. We add elements to this initial ROB following a similar philosophy to that of Algorithm 1, in which we generate N_{lim} random waveforms, compute

Algorithm 6. Enrich ROB to construct an EIM under tolerance.

-
-
- 1: **Input:** Initial ROB $\{\vec{e}_i\}_{i=1}^n$ and EIM $\{\beta_i\}_{i=1}^n$, maximum number of waveforms selected N , tolerance σ , maximum number of waveforms computed N_{lim} , maximum number of eigenvectors used n_λ
 - 2: **repeat**
 - 3: Generate $N_{\text{lim},j}$ waveforms $\{h_A\}_{A=1}^{N_{\text{lim},j}}$ and compute their EIM error $\sigma_{\text{EIM},A}$
 - 4: Select the N waveforms $\{h_A\}_{A=1}^N$ with largest σ_{EIM}
 - 5: Save the minimum value of σ_{EIM} for the selected waveforms: $\sigma_{\text{EIM},\min}$
 - 6: $M_{AB}^{\text{ROB}} = \langle h_A, h_B \rangle - \sum_{c=1}^{n_\lambda-1} \langle h_A, e_c \rangle \langle e_c, h_B \rangle$
 - 7: Diagonalize M_{AB}^{ROB} and obtain eigenvalues λ_A and eigenvectors E_{AB}
 - 8: Compute the n_λ normalized eigenvectors in waveform domain with largest $\delta\sigma_{A,\max} = \lambda_A \max_B |E_{AB}|^2$: $\{\vec{u}_A\}_{A=1}^{n_\lambda}$
 - 9: **repeat**
 - 10: Compute the maximum contribution of each eigenvector to $\sigma_{\text{EIM},A}$:
 $\{\delta\sigma_{A,\max}^{\text{EIM}} = \lambda_A (1 + \|(\hat{P}^\dagger \hat{V})^{-1} \hat{P}^\dagger \vec{u}_B\|^2) \max_B |E_{BA}|^2\}_{A=1}^{n_\lambda}$
 - 11: Find largest $\delta\sigma_{A,\max}^{\text{EIM}}$: $A_{\text{sel}} = \text{argmax}_A (\delta\sigma_{A,\max}^{\text{EIM}})$
 - 12: Add the corresponding eigenvector to the ROB:
 $\{\vec{e}_i\}_{i=1}^n \leftarrow \{\vec{e}_i\}_{i=1}^n \cup \{\vec{u}_{A_{\text{sel}}}\}$
 - 13: Remove the selected eigenvector from the eigenvector list: $\{\vec{u}_A\}_{A=1}^{n_\lambda} \leftarrow \{\vec{u}_A\}_{A=1}^{n_\lambda} \setminus \{\vec{u}_{A_{\text{sel}}}\}$
 - 14: Input $\{\vec{e}_i\}_{i=1}^n$, N_{rounds} , $\{\vec{u}_A\}_{A=1}^{n_\lambda}$ and their corresponding eigenvalues $\{\lambda_A\}_{A=1}^{n_\lambda}$ into Algorithm 5 to obtain a new EIM $\{\beta_i\}_{i=1}^n$.
 - 15: Find new error of selected waveforms $\{\sigma_{\text{EIM},A}^{\text{new}}\}_{A=1}^N$
 - 16: **until** $\max_A \sigma_{\text{EIM},A}^{\text{new}} \leq \sigma$
 - 17: **until** $\sigma_{\text{EIM},\min} < \sigma$
 - 18: **Output:** ROB $\{\vec{e}_i\}_{i=1}^n$
-
-

their EIM error σ_{EIM} , and select the N waveforms with largest EIM error for further study. Again, we want N to be as large as allowed by the memory [see Eq. (21)]. We then compute the matrix M_{AB}^{ROB} for the N selected waveforms, find its eigenvalues λ_B and compute the $n_\lambda < N$ most relevant eigenvectors in the waveform domain $\{\vec{u}_B\}_{B=1}^{n_\lambda}$, where the value of n_λ is again limited by the memory of the system. We iteratively select the eigenvector with the largest contribution to the EIM error, add it to the ROB, and construct a new EIM with Algorithm 5 until all N waveforms are fitted with a tolerance better than the required one. The process is summarized in Algorithm 6.

IV. CODE VALIDATION

In this section, we aim to quantify and assess the validity of the ROQ basis obtained using the algorithm described in Sec. III. For that matter, we would like to evaluate the accuracy of the different bases in reconstructing the original waveform as well as the speedup gained. First, in Sec. IV A we describe the bases to be tested and compare them with examples found in the literature, in Sec. IV B we show the results of two statistical tests for the various bases, in Sec. IV C we comment on the theoretical and empirical speedups using the ROQ, and finally in Sec. IV D we compare the results of doing a parameter estimation analysis with the standard and the ROQ likelihoods.

A. Basis generation and comparison with other ROQ methods

In this section, we describe how we generate the bases that will be used for testing and parameter estimation. We construct bases for both IMRPhenomPv2 [27] and IMRPhenomXPHM [28]. Both waveform models take into account the effects of spin precession and IMRPhenomXPHM also includes higher order mode GW emission.

For IMRPhenomPv2 we generate the bases listed in Table I, covering a chirp mass (\mathcal{M}) range between $0.95M_\odot$ and $45M_\odot$. Given that integration is performed from a low-frequency cutoff of 20 Hz, we find the bases' duration ranging from 256 to 4 s. For IMRPhenomXPHM, we generate the bases listed in Table I, with chirp masses ranging between $2.18M_\odot$ and $110M_\odot$ and corresponding durations between 64 and 4 s from 20 Hz. We show in Fig. 2 an example of an IMRPhenomXPHM waveform and its corresponding empirical interpolant. More specifically, the upper panel shows the real part of the plus polarization $\text{Re}(\tilde{h}_+)$ and the lower panel shows its square $|\tilde{h}_+|^2$ in the frequency domain. The corresponding interpolation nodes and empirical interpolant are shown to visually confirm the goodness of the fit to the original waveform. The parameters of the template are shown in the caption of Fig. 2 and are selected so that the quadratic EIM error is equal to the median quadratic EIM error over the testing set of waveforms of the basis of Table II covering $\mathcal{M} \in [10, 15]M_\odot$.

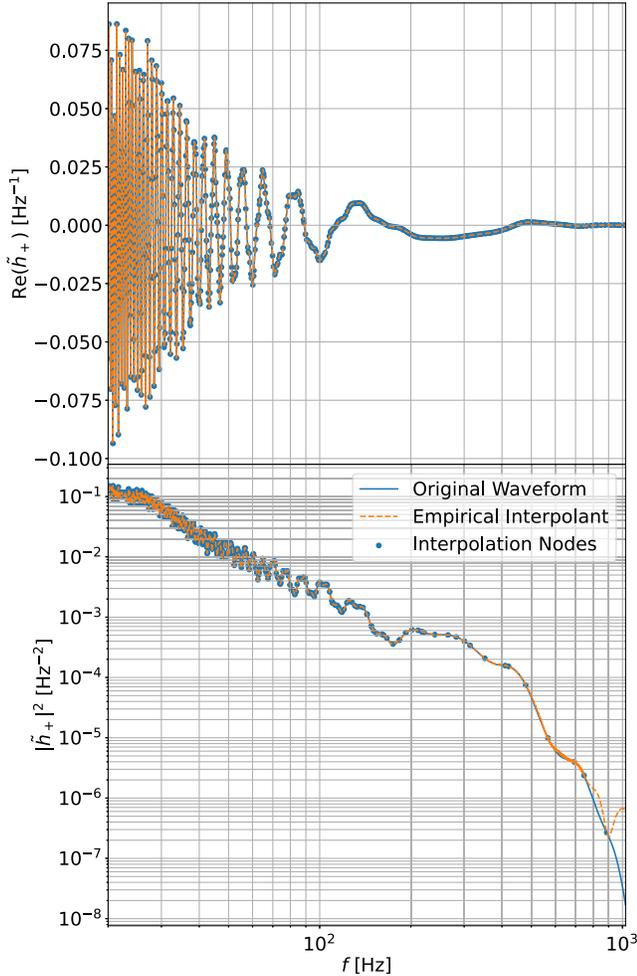


FIG. 2. Example of an IMRPhenomXPHM template and its empirical interpolant. Upper: we show the real part of the plus polarization of the template $\text{Re}(\tilde{h}_+)$ as a function of the frequency. Lower: we depict its square $|\tilde{h}_+|^2$. We superimpose in each panel the corresponding interpolation nodes and empirical interpolants as defined in (24). The template shown has $\mathcal{M} = 13.6M_\odot$, $q = 2.61$, $\chi_{\text{eff}} = -0.011$ [38,39], $\chi_p = 0.208$ [40], and inclination angle $\iota = 61.6^\circ$. Using the ROQ basis of Table II covering $\mathcal{M} \in [10, 15]M_\odot$, we have linear and quadratic EIM errors of $\sigma_{\text{linear}}^{\text{EIM}} = 1.26 \times 10^{-9}$ and $\sigma_{\text{linear}}^{\text{EIM}} = 6.06 \times 10^{-8}$, respectively.

We can observe how both the linear and quadratic parts have a complicated dependence on frequency, coming from the interference of the higher order modes with the main (2,2) mode. This is the principal reason for the larger number of linear and specially quadratic elements when comparing the basis of IMRPhenomPv2 and IMRPhenomXPHM.

The 4 and 8 s basis of IMRPhenomPv2 and IMRPhenomXPHM are directly comparable with those published in Ref. [21] computed using PyROQ, since they cover the exact same parameter space and frequency range. We observe that the number of basis elements in PyROQ and EigROQ is generally similar and we expect it to be smaller

than that of a comparable basis constructed with GreedyCPP. However, the number of test points over the set tolerance is about an order of magnitude smaller in our bases than in the PyROQ ones.⁴ We attribute this improvement to the way we approach the minimization in the error of the empirical interpolant. In the PyROQ algorithm, it is implicitly assumed that once a template is below the tolerance it will remain like this throughout the computation, which would be true if the EIM error were monotonically decreasing. This, however, is not true, in general, as adding new templates to the base can deteriorate the fit and, in particular, it can bring some of the waveforms that were under the tolerance, back over tolerance. The fact that this is happening can be explicitly seen in Ref. [21] because the maximum EIM error in the training set is over the tolerance. To alleviate this problem, we simultaneously use the N waveforms with initially more EIM error, even if some of them are already below tolerance.

We have also extended the parameter space of the ROQ bases with respect to those computed by PyROQ in Ref. [21], with durations up to 256 s for IMRPhenomPv2 and 64 s for IMRPhenomXPHM. Doing this in PyROQ is computationally challenging since finding the template with the largest associated EIM error requires the recomputation of the waveforms in the training set many times. With our methods, this is no longer the case, as we only need to compute any given waveform once. This allows more complex case studies to be feasible.

B. Statistical tests

In this section, we perform two different statistical tests to check the faithfulness of the ROQ basis in gravitational waves inference, a likelihood test and a percent-percent (P-P) test.

The likelihood test consists of a comparison of the log-likelihood ratios evaluated using the standard waveform with those obtained using the ROQ approximation. The log-likelihood ratio is defined as the ratio between the likelihood of Eq. (2) and the likelihood of the noise hypothesis ($h = 0$); that is,

$$\begin{aligned} \log \mathcal{L}_{\text{ratio}}(d|\vec{\theta}) &= \log \frac{\mathcal{L}(d|h(\vec{\theta}))}{\mathcal{L}(d|0)} \\ &= (d, h(\vec{\theta})) - \frac{1}{2}(h(\vec{\theta}), h(\vec{\theta})). \end{aligned} \quad (54)$$

This quantity, which is just the likelihood of Eq. (2) removing the constant part that only depends on the data, is what we will be referring to as the log-likelihood throughout the rest of the text. The log-likelihood is the crucial quantity used in estimating the parameters of a given GW event,

⁴Note that while in Ref. [21] the bases use 10^6 points for testing, we use 10^7 points.

which is the ultimate end for which the ROQ is created. We perform likelihood tests on the IMRPhenomPv2 and IMRPhenomXPHM bases described in Tables I and II, respectively, and show the results in Fig. 3. To obtain the difference in the log-likelihood, we create a random realization of Gaussian noise and inject a waveform calculated using the corresponding approximant. The injected waveforms' parameters are randomly sampled from uniform distributions whose boundaries are the respective ROQs' ranges of validity. We use a fixed distance of 100 Mpc and randomly sample the incoming direction of the GW from a uniform distribution in the sky. We then compute the standard log-likelihood and the ROQ log-likelihood using the same injected waveform and compare them. What we plot is the relative difference between both logarithms for a total of 1.5×10^5 realizations. We see the maximum discrepancy lies below 0.1 for every case considered here, and the bulk of the samples lie below 10^{-3} .

Given that the likelihood is the only signal-dependent quantity that enters the computation of the posterior [Eq. (1)], as long as the ROQ and standard likelihoods agree

reasonably well, we can expect the PE posteriors with and without the ROQ to be virtually the same. According to the Wilks theorem [41] in the frequentist and large sample size limits, the quantity $-2 \log\{\mathcal{L}/\mathcal{L}_{\max}\}$ is distributed as a χ^2 with a number of degrees of freedom equal to the number of parameters being fitted by the PE. In the case of a CBC, we need 15 parameters to fully characterize the binary, although, since the azimuthal spin angles and phase of coalescence are usually so poorly constrained, in most cases the effective number of parameters is reduced to 12. We then expect $\log\{\mathcal{L}/\mathcal{L}_{\max}\} = -5.7_{-4.8}^{+3.1}$, which is in accordance with most of the GW observations, especially those with high signal-to-noise ratio (SNR). Under the same model, the standard deviation of $\log \mathcal{L}$ is $\sigma_{\mathcal{L}} \sim \sqrt{N_{\text{eff}}}/2 \sim \sqrt{6} \sim 2.4$, where N_{eff} is the effective number of parameters. Therefore, as long as the difference between the logarithm of the standard and the ROQ likelihoods is much smaller than $\sigma_{\mathcal{L}} \sim 2.4$, we expect the posteriors to be similar.

From Eq. (54), we observe that the likelihood ratio of a GW signal will approximately be given by $\log \mathcal{L} \sim \rho^2/2$, where ρ is the SNR. Therefore, the condition that $\Delta \log \mathcal{L} \ll 2.3$ can be translated into a condition on the SNR,

$$\rho \ll 2 \sqrt{\frac{\log \mathcal{L}}{\Delta \log \mathcal{L}}}, \quad (55)$$

which can be used to interpret Fig. 3 in terms of up to which SNR we can trust the posteriors obtained when using the corresponding ROQ. If we want the ROQ to be valid for the analysis of larger SNRs, we can always decrease the tolerance σ with which we generate it, at the expense of having more basis elements.

The second of the tests is the percent-percent plot [42,43]. P-P plots have been widely used in the literature [44] to validate codes that perform Bayesian parameter estimation. Therefore, we use the P-P plots to directly test the ROQ's faithfulness in its intended use. In this specific case, to make the P-P plots shown in Fig. 4, we use the posteriors probability density functions (PDFs) resulting from performing PE on 200 injections. The PE analyses are done using the ROQ likelihood and the DYNESTY [45] sampler within the BILBY [46] framework, and the injections use the same waveform model for which the corresponding ROQ was constructed. The priors of the PE and the distribution from which the injections are drawn are the same and coincide with the parameter space in which each ROQ basis has been constructed. For the extrinsic parameters, we put priors that are uniform in the sky and in comoving volume, going to a maximum distance tailored for each chirp mass range to have detectable signals.

In the P-P plots of Fig. 4 we show the fraction of posterior PDFs for which the injected value of the parameter is found in a given confidence interval (C.I.) as a function of that same confidence interval. We expect the fraction of injected parameter values that fall into a

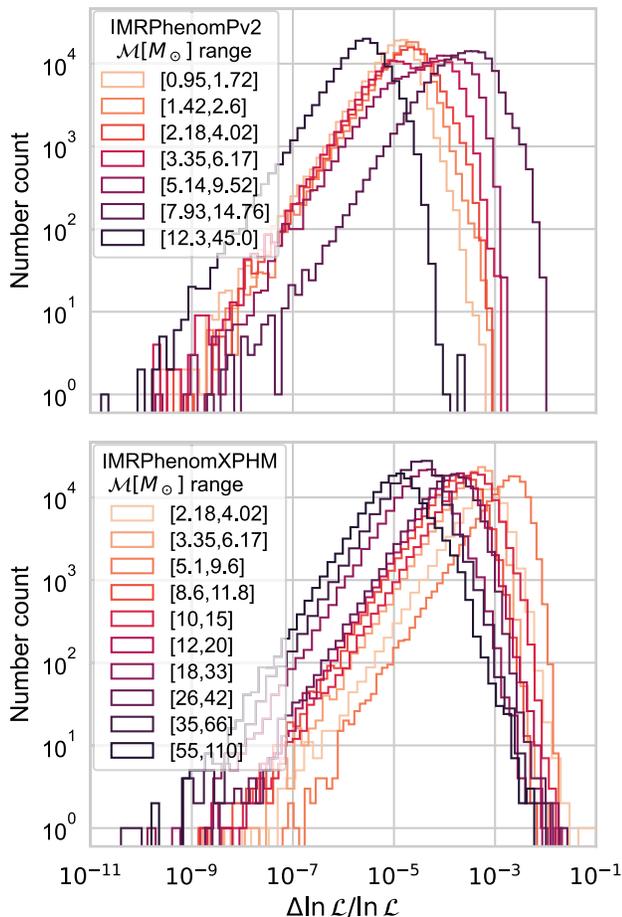


FIG. 3. Likelihood error tests for various \mathcal{M} ranges. Specifically, we plot $\Delta \ln \mathcal{L} / \ln \mathcal{L}$, that is, the fractional error of the $\ln \mathcal{L}$ when calculated with and without the ROQ. Upper: IMRPhenomPv2. Lower: IMRPhenomXPHM.

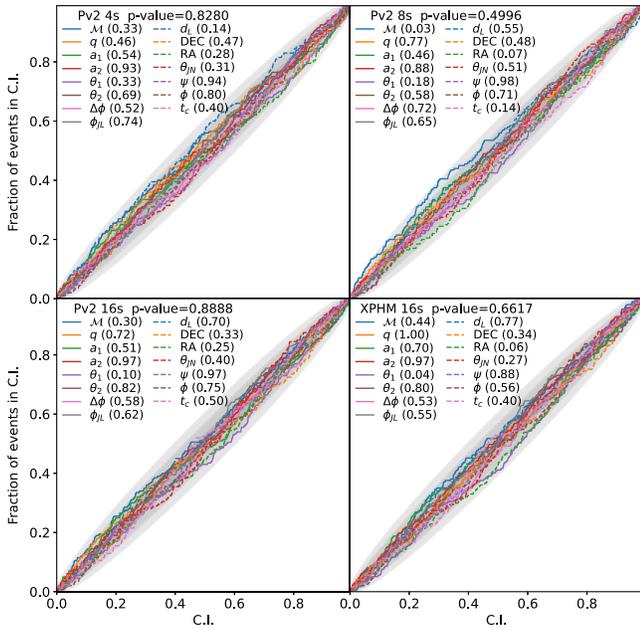


FIG. 4. P-P plots performed with different ROQ bases as stated in each subplot’s legend. We show here the result of 200 injections being drawn from the corresponding ROQ-compatible prior, as stated in Tables I and II. The contours in gray delimit the 1σ , 2σ , and 3σ regions. We plot a line for every parameter that uniquely characterizes a given CBC with consistent colors and styles across subplots. The lines represent the cumulative fraction of events.

particular confidence interval of the posterior PDFs to be drawn from a uniform distribution. We can thus assign a p value quantifying such claim [43], individually for each binary parameter and jointly for all the parameters. The p values are shown in the legends of Fig. 4. For all the P-P plots shown, the cumulative distribution functions of the majority of the parameters fall within the 3σ regions, leading to p values that are consistent with a uniform distribution. The combined p values lie between 0.49 and 0.89, indicating that the posterior PDFs produced using these ROQs are well calibrated.

C. Speedup analysis

The main purpose of the ROQ is to accelerate the computation of the GW likelihood. To test how good it is in this regard we perform a series of speedup trials shown in Fig. 5. There are two quantities that we evaluate for the benchmarking test: the waveform and the Gaussian log-likelihood described in Eq. (2). The tests consist of timing several calculations of both quantities for the standard case and the ROQ case. The sets of parameters used as inputs are drawn from uniform distributions with boundaries based on the range of validity of the corresponding ROQ basis. The ratio between the time for the standard method and the ROQ is what we call the empirical speedup, where we use the term empirical because we perform the actual likelihood

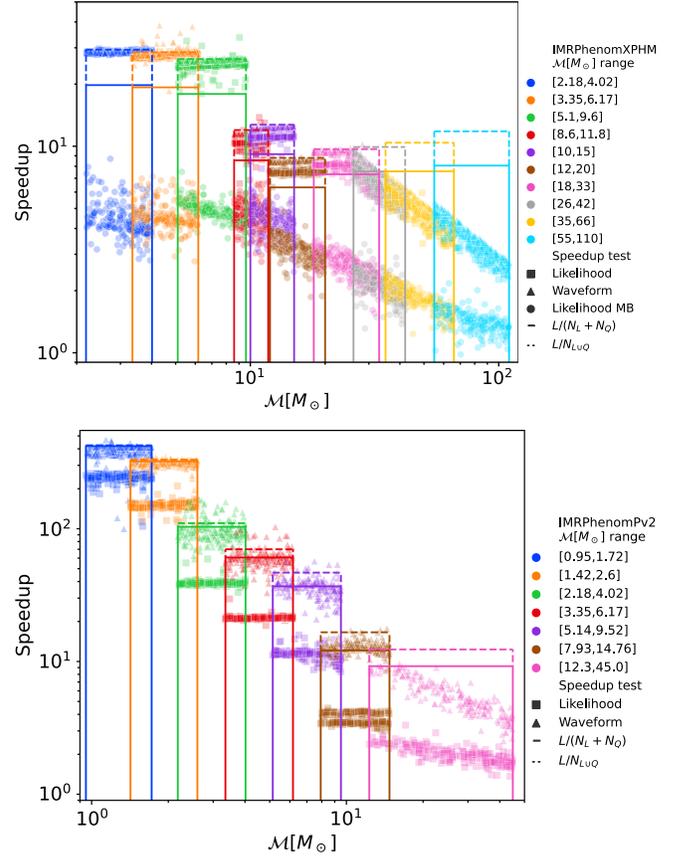


FIG. 5. Speedup factor for the IMRPhenomXPHM (upper) and IMRPhenomPv2 (lower) waveforms in the different regions in chirp mass where the ROQ has been computed. We can differentiate theoretical and empirical speedups. The empirical speedups are calculated as the ratio between the time spent in computing the waveform without and with the ROQ and are plotted as triangles. Squares are obtained in the same way, but employing the likelihood. In the IMRPhenomXPHM case, we include the likelihood speedups with the multibanding option enabled (circles) and disabled (squares). The theoretical speedups are drawn as bars. The dashed bars represent the speedup when array frequency duplications are accounted for, while solid bars are for when they do not.

and waveform computations using PYTHON [47] and the BILBY [46] framework. For IMRPhenomXPHM waveform speedups, we disable the default multibanding [37], which is used to speed up the full waveform computation by reducing the number of frequencies the model is evaluated at and then interpolates between them. Therefore, we disable this to test if the model is linear with the number of frequencies at which it is evaluated. However, for the likelihood test, we compute the speedups both without and with multibanding enabled, to explore real-world speedup gains.

In Fig. 5, we differentiate the speedups using triangles for the waveform, squares for the log-likelihood, and in the IMRPhenomXPHM case, circles for the log-likelihood with multibanding enabled. We can also compare with the theoretical speedups that are plotted as histogramlike

bars in the figure. There are two kinds of bars, the solid ones represent the quantity

$$\text{theoretical speedup} = \frac{L}{N_L + N_Q}, \quad (56)$$

where L is the number of frequencies for the waveform evaluation in the standard computation, and N_L and N_Q are the frequency nodes for the linear and quadratic ROQ bases without factoring out repeated frequencies. This is the theoretical speedup that is usually attributed to the ROQ in the literature [20]. The dashed bars are the same quantity as in Eq. (56) when the frequencies belonging to both the linear and the quadratic interpolation list of frequency nodes are just considered once, thus the notation L/N_{LUQ} . In the ROQ likelihood we need to call the waveform model only once at the frequencies defined by $\{f_i\}_{i=1}^{N_{\text{LUQ}}} = \{F_j\}_{j=1}^{N_L} \cup \{\mathcal{F}_k\}_{k=1}^{N_Q}$, as is done in BILBY. Therefore, L/N_{LUQ} will be the theoretical speedup of the waveform evaluation if we assume that its computation time is proportional to the number of sampling points. For the IMRPhenomXPHM case, the difference between $N_L + N_Q$ and N_{LUQ} can be significant since there are many repeated interpolation nodes at low frequencies. The reason is that, in the low-frequency region, the amplitude is larger and the waveform oscillates more rapidly than in the high-frequency part. Consequently, the interpolation nodes tend to concentrate at low frequencies, leading some of them to coincide in the linear and quadratic ROQ. This behavior can be seen in Fig. 2.

For \mathcal{M} smaller than $\sim 20M_\odot$, we see that the waveform speedups are constant in the entire \mathcal{M} range of a given basis and are always close to the theoretical value of L/N_{LUQ} . This is in agreement with our expectations, since the IMRPhenom models describe the inspiral in a way that the computation time is linear with the number of sampling points, and their implementation in LALSimulation [48] being tested is written efficiently in C [49], with minimal overheads. In the case of large \mathcal{M} , above $\sim 20M_\odot$, the

waveforms start being dominated by the merger and ring-down, the last two phases of a CBC, which are harder to model, and the speedup of the IMRPhenom models is smaller than the theoretical expectation. This can be due to the waveform generation stopping above the ringdown frequency, meaning that the model is evaluated at fewer frequency points for high mass signals. Furthermore, when the waveform uses sufficiently few frequency points, fixed costs associated with calculating post-Newtonian and phenomenological parameters of the model become important. Therefore, as \mathcal{M} increases, the trend of the waveform speedup is to decrease until a value of $\mathcal{O}(1)$ is reached and we have no speedup at all.

For the IMRPhenomXPHM likelihood speedups, we show both the results with and without disabling the default multibanding [37], which is used in the standard likelihood to speed up the full waveform computation. We observe that without multibanding IMRPhenomXPHM has a likelihood speedup very close to the theoretical value. This is due to the fact that the computation time of the likelihood is dominated by the waveform evaluation, and the BILBY implementation of the ROQ likelihood only generates the waveform once at the frequencies $\{f_i\}_{i=1}^{N_{\text{LUQ}}} = \{F_j\}_{j=1}^{N_L} \cup \{\mathcal{F}_k\}_{k=1}^{N_Q}$. However, when one includes the multibanding option, the IMRPhenomXPHM is already internally being evaluated in fewer frequency points, and therefore the speedup can be significantly lower than the expected one, although it still reaches median values that can be as large as 5 and which will be noticeable in PE applications. Looking at the targeted bases that are introduced in Table III, we observe that, in this case, the speedup over the standard multiband case can be even larger, reaching a value of $29.2_{-4.6}^{+1.4}$ for the base targeted at GW170817 [50].

In the IMRPhenomPv2 case, we observe that the likelihood speedup is significantly below the waveform speedup and, therefore, also below the theoretical speedup. To understand this discrepancy, we note that, for the standard likelihood case, the computation time is dominated by evaluating the waveform in all the required frequencies

TABLE III. Focused IMRPhenomXPHM bases for GW190814 ($\Delta f = 1/16$ Hz) and GW170817 ($\Delta f = 1/256$ Hz). We limit the magnitudes of the two spins $-0.8 \leq \chi_i \leq 0.8$ for $i \in [1, 2]$ and the full range for the spin angles $(0, 0) \leq (\theta_j, \alpha_0) \leq (\pi, 2\pi)$. For the GW190814, we limit the mass ratio $q \leq 16$, while for GW170817 we limit it $q \leq 4$. For the creation of the two bases, we run EigROQ with the same configuration. In Algorithm 1, we set the maximum number of waveforms selected $N = 20000$, tolerances of each step $\sigma_i = [10^{-2}, 10^{-3}, 10^{-4}]$, and maximum number of waveforms computed per step $N_{\text{lim},i} = [10^5, 3.16 \times 10^5]$. In Algorithm 6, we set $N = 10^6$, $\sigma = 10^{-4}$, $N_{\text{lim}} = 10^7$, and the maximum number of eigenvectors used $n_\lambda = 5000$. The theoretical speedup has been computed with Eq. (56), while the empirical speedup is the median and 90% credible interval of the corresponding points in the lower panel of Fig. 5. For the empirical speedups, we show the values both without (Emp.) and with (MB) the IMRPhenomXPHM multibanding option enabled [37].

Frequency range (Hz)			$M_c(M_\odot)$		Basis size		Likelihood speedup		
Min.	Max.	Δf (Hz)	Min.	Max.	Linear	Quadratic	Theoretical	Empirical	MB
20	2048	1/16	6.2	6.6	1090	816	17.0	$21.6_{-3.6}^{+3.8}$	$4.8_{-0.6}^{+1.1}$
20	2048	1/256	1.195	1.200	1392	2007	152.7	$151.8_{-4.1}^{+4.5}$	$29.2_{-4.6}^{+1.4}$

and computing the overlap integrals of Eq. (2), both of which will be proportional to L . However, for the ROQ likelihood, the time to compute the waveform and overlap integrals is significantly reduced since they are proportional to $N_L + N_Q \ll L$. Given the fact that IMRPhenomPv2 is much faster to generate than IMRPhenomXPHM, the computation time starts to be dominated by fixed-cost operations, which, for example, include computing the parameters of the waveform models and finding the detector responses as well as possible overheads.

To further explore this hypothesis, we model the computation time of the likelihood as a coefficient multiplying the number of frequencies being evaluated plus a constant term that represents the fixed-cost operations. Since for IMRPhenomPv2, $N_{L+Q} \sim N_L + N_Q$, we have

$$T = A \cdot L + B, \quad (57)$$

$$T_{ROQ} = a \cdot (N_L + N_Q) + b. \quad (58)$$

To compute the speedup, we divide Eq. (57) by Eq. (58), obtaining

$$f(L, N_L, N_Q; B, a, b) = \frac{L + B}{a(N_L + N_Q) + b}, \quad (59)$$

where we have divided all the coefficients by A , which is not expected to be 0. In Fig. 6, we show the ratio between the empirical and theoretical likelihood speedups, together with the best fit of our model in Eq. (59). We observe very good agreement between the model and the data. From the fitted values of B , a , and b , also displayed in the plot, we can substantiate our hypothesis that the fixed-cost operations in

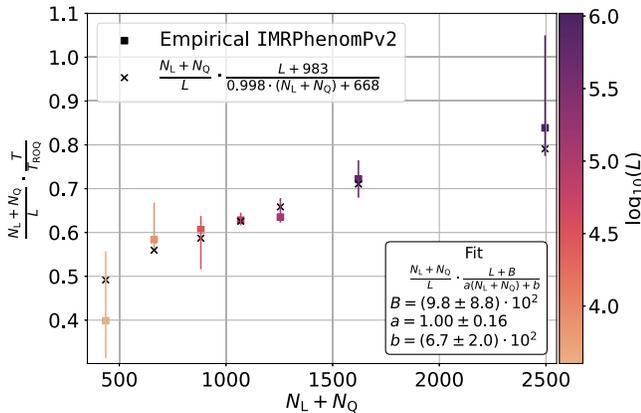


FIG. 6. Ratio between empirical speedup and the theoretical speedup of Eq. (56), plotted as a function of the total elements of the ROQ basis ($N_L + N_Q$) for IMRPhenomPv2. The color of the error bars encodes the logarithm of the number of frequencies where the waveform is evaluated in the standard computation $\log \mathcal{L}$. In the inset, we show the functional form we fit, which comes from Eq. (59), as well as the 1σ uncertainty for the three fitted parameters. We also plot as black crosses the results obtained evaluating the best fit in the data points.

the ROQ likelihood is making the empirical speedup of the IMRPhenomPv2 smaller than the theoretical value. We find a value of $a = 1.00 \pm 0.16$, and therefore, from Eq. (59), we observe that, if the coefficients B and b describing the fixed costs were zero, we would recover the theoretical speedup result. However, since we find a value of $b = (6.7 \pm 2.0) \times 10^2$, the IMRPhenomPv2 speedup is reduced with respect to the theoretical unless we have a very large number of basis elements such that $a \cdot (N_L + N_Q) \gg b$.

D. Application to GW events

We now perform four PE analyses [10] on three confirmed GW events using the ROQ approximation. More specifically, we use the IMRPhenomXPHM 16 s basis described in Table II for the GW191129_134029 [51] event and the IMRPhenomPv2 256 s basis of Table I for the GW170817 [50] event. For the other two PE analyses of GW190814 [52] and GW170817 with IMRPhenomXPHM, in a similar spirit to Refs. [22,23], we construct targeted ROQ bases with narrow \mathcal{M} ranges, listed in Table III. These bases are centered on the search \mathcal{M} value and have a narrow width tuned to be larger than the expected chirp mass resolution. Note that the bases have been generated using a factor of 10 times fewer waveforms than that of Tables I and II, since the parameter space they cover is smaller.

The analyses use the ROQ likelihood and the DYNESTY [45] sampler within version 2.1.0 of BILBY [46] and the version 5.1.0. of LALSimulation. The PSDs employed were estimated using BayesWave [53,54] and are those used by the LVK Collaboration for the public analysis of the events. We also include the effects of calibration uncertainties [55–57] in the phase and the amplitude.

The first event we discuss is GW191129_134029 [51,58]. This is an event with $\mathcal{M}^{\text{detector}} = 8.48^{+0.06}_{-0.05} M_\odot$, so we can use the 16 s IMRPhenomXPHM ROQ basis. It has a relatively big median network SNR of 13.1, allowing us to put tight constraints on the parameters and better see if any differences arise between the ROQ and the standard posterior. We perform two BILBY runs with the exact same configuration, one using the standard GW likelihood and the other using the ROQ likelihood.

In Fig. 7, we show the difference between the logarithm of the standard and the ROQ likelihoods, for the posterior samples of the PE with the ROQ likelihood. This difference corresponds to the ROQ error in the log-likelihood. We find a 90% C.L. error of $\Delta \log \mathcal{L} = 0.075^{+0.051}_{-0.057}$. Since $\Delta \log \mathcal{L} \ll 1$, we expect the posteriors with and without the ROQ to be almost the same. Using that the log-likelihood of this event is $\log \mathcal{L} = 84.2^{+2.9}_{-4.1}$, and the fractional error in the ROQ log-likelihood computation is $\delta_{\mathcal{L}} = (9.1^{+6.3}_{-6.8}) \times 10^{-4}$.⁵ The distribution of errors is

⁵We define the fractional error in the ROQ log-likelihood computation as $\delta_{\mathcal{L}} = \Delta \log \mathcal{L} / \log \mathcal{L}$.

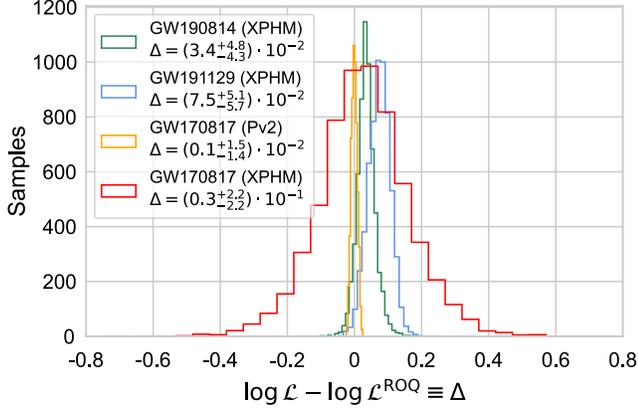


FIG. 7. Difference between the logarithm of the standard likelihood and the logarithm of the ROQ likelihood for the three events analyzed.

centered at a positive value, as one would expect if the waveform model were a good representation of reality, since any error in the ROQ modelization of the waveform would push it away from the true GW and, thus, to a lower likelihood value.

In Fig. 8, we corroborate that, indeed, the posteriors with and without the ROQ are similar by showing the corresponding distributions for the detector frame chirp mass \mathcal{M} and the mass ratio q . We find a Jensen-Shannon divergence [59] of 1.3×10^{-4} and 1.9×10^{-4} , respectively,

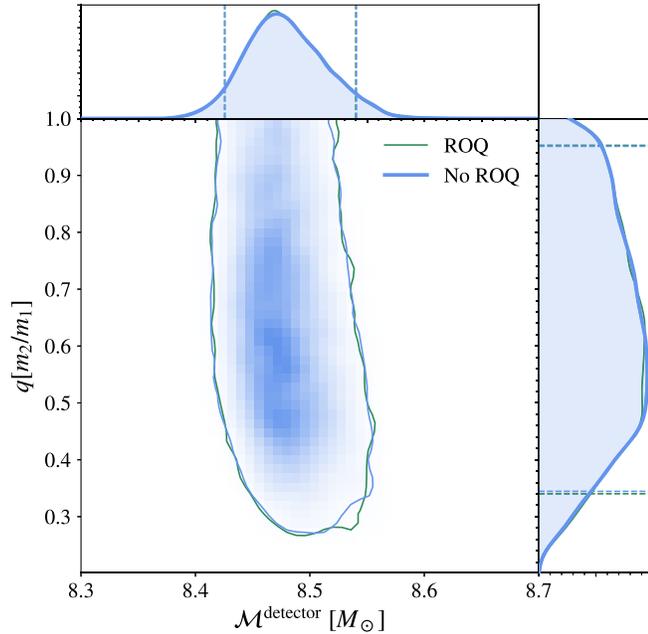


FIG. 8. Posterior distributions for the mass ratio and \mathcal{M} in the detector frame for the ROQ and non-ROQ analysis of GW191129_134029. The 90% credible regions are indicated by the solid contour in the joint distribution and by the dashed vertical and horizontal lines in the marginalized distributions.

robustly assessing the similarity between the distributions with and without the ROQ approximation.

The second event we analyze is GW190814 [52,58]. This event was measured to have a chirp mass of $\mathcal{M} = 6.42^{+0.02}_{-0.02} M_{\odot}$ and a very unequal mass ratio of $0.11^{+0.01}_{-0.01}$, which is below the mass ratios of $q > 0.25$ explored in the bases of Table II. Therefore, we create a targeted ROQ base with 16 s of duration and a chirp mass range from $6.2 M_{\odot}$ to $6.6 M_{\odot}$ for the IMRPhenomXPHM waveform. In Fig. 7, we show the ROQ log-likelihood errors of the posterior samples of the PE performed using this targeted basis. We have that $\Delta \log \mathcal{L} = 0.034^{+0.048}_{-0.043}$, which is similar in magnitude to that of GW191129_134029. Again, since $\Delta \log \mathcal{L} \ll 1$, we expect the posteriors with and without the ROQ to be almost the same. However, for this event, the log-likelihood is larger, at $\log \mathcal{L} = 310.3e + 02^{+3.1}_{-5.0}$, and therefore the relative error in the ROQ log-likelihood computation is smaller, at $\delta_{\mathcal{L}} = (1.1^{+1.5}_{-1.4}) \times 10^{-4}$.

The last GW event we analyze is GW170817 [50], the event with the largest network SNR (~ 33) ever detected. It was identified as a binary neutron star with $\mathcal{M} = 1.1976^{+0.0004}_{-0.0002} M_{\odot}$ [60] and we use it to probe the longest of

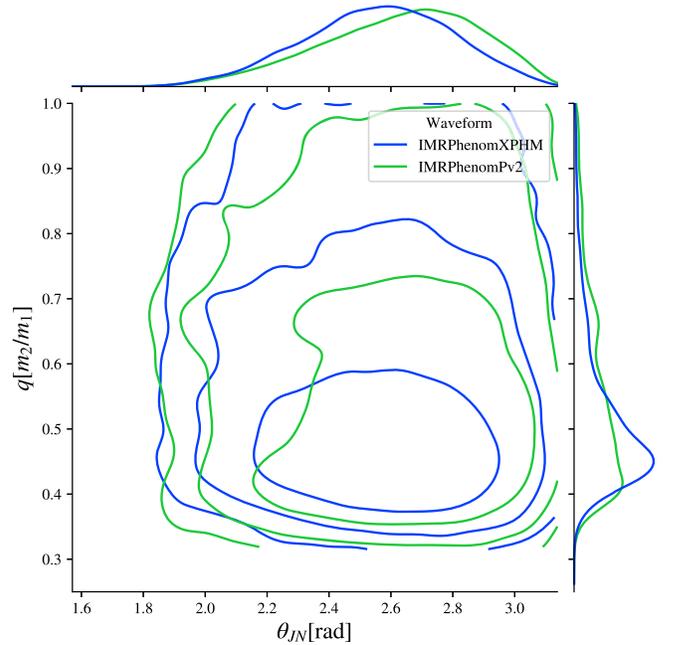


FIG. 9. Posterior distributions for the mass ratio q and the inclination angle θ_{JN} for the ROQ analysis of GW170817. In blue we plot the IMRPhenomXPHM run and in green IMRPhenomPv2. Three contours per run delimit the 1σ (68.3% C.L.), 2σ (95.4% C.L.), and 3σ (99.7% C.L.) credible regions in the joint $q - \theta_{JN}$ distribution. Note that the noncontinuous behavior of the contours near the border is an artifact of the Gaussian kernel employed in the drawing. This is expected whenever the parameter is bounded and presents many samples close to the border.

our IMRPhenomPv2 bases with 256 s in duration as well as a targeted ROQ using IMRPhenomXPHM for such long signals. For our analysis, we make use of the public strain data after noise subtraction [61]. In Fig. 7, we show the ROQ log-likelihood errors of the posterior samples of both PEs. For both cases, we do not expect the ROQ error to significantly impact the posterior, since $\Delta \log \mathcal{L} \ll 2.3$. The IMRPhenomPv2 PE has an order of magnitude smaller ROQ error than the IMRPhenomXPHM case. This is most likely the result of the IMRPhenomPv2 basis being constructed with a tolerance $\sigma = 10^{-5}$, which is an order of magnitude smaller than the tolerance $\sigma = 10^{-4}$ used in the IMRPhenomXPHM case. In the IMRPhenomPv2 case, the log-likelihood is $536.1_{-4.3}^{+3.2}$ and the corresponding fractional error is $\delta_{\mathcal{L}} = (-0.1_{-2.6}^{+2.8}) \times 10^{-5}$. In the IMRPhenomXPHM case, we find a larger likelihood of $538.1_{-5.1}^{+4.3}$, which is expected since the higher order modes give more freedom to the waveform to fit the data. The corresponding fractional error is $\delta_{\mathcal{L}} = (0.5_{-4.0}^{+4.1}) \times 10^{-4}$. Comparing the Bayes factors of both PE runs, adjusted to have the same priors, we find $\log \mathcal{B} = 1.1 \pm 0.3$ in favor of IMRPhenomXPHM, which can be taken as evidence for higher order modes in the signal. This highlights the importance of considering all physical effects of the waveform. To further make this point, we show in Fig. 9 how the addition of the higher order modes improves the determination of the mass ratio and the inclination angle θ_{JN} , even for this low mass CBC for which the higher order modes are harder to measure in LIGO-Virgo [12].

V. CONCLUSIONS

In this paper, we have explored in-depth reduced order quadrature methods applied to GW data analysis and have presented novel algorithms to improve different aspects of the ROQ bases' construction. ROQ methods offer a significant advantage by reducing the computational burden associated with likelihood evaluations, especially for long-duration waveforms and therefore can greatly speed up parameter estimation analyses. Existing procedures for constructing ROQ bases encounter challenges in approximating waveforms that include complicated features such as precession or higher order modes. We present algorithms to address these limitations by making use of SVD methods to characterize the waveform space and choose a reduced order basis close to optimal. We also propose improved methods to select the empirical interpolation nodes, greatly reducing the error induced by the empirical interpolation model.

We have demonstrated the effectiveness of our algorithm by constructing multiple ROQ bases for the IMRPhenomPv2 and IMRPhenomXPHM waveforms, ranging in duration from 4 to 256 s. These bases have been subjected to various tests, including likelihood error tests and P-P tests, to validate their accuracy and trustworthiness for data analysis applications. The speedup of

these bases has also been empirically explored, confirming that ROQ methods provide close to the expected reduction in computational time compared to traditional likelihood calculations.

Furthermore, we have performed end-to-end parameter estimation analyses on several confirmed GW events. The results provide compelling evidence of the algorithm's ability to generate ROQ bases that accurately represent complex waveform models over both broad and targeted parameter spaces. By directly comparing the posterior distributions using the ROQ and standard methods and understanding the log-likelihood error distributions, we validate that our bases can straightforwardly be incorporated into current pipelines to produce precise and unbiased parameter estimations in real gravitational wave detector data.

In conclusion, the algorithms introduced in this paper represent a step forward in the quest to efficiently exploit the capabilities of advanced gravitational wave detectors. We improve upon previous ROQ construction algorithms allowing for more efficient bases in regions of parameter space that were previously inaccessible. As gravitational wave astronomy continues to evolve, and the number of events detected per year continues to grow, having fast and accurate techniques to perform parameter estimation will undoubtedly play a vital role in maximizing the scientific potential of future observatories and advancing our knowledge of the Universe.

The authors acknowledge the use of the publicly available codes: BILBY [46], LALSuite [48], and EigROQ [29]. This research has made use of data or software obtained from the Gravitational Wave Open Science Center [62], see also [58,63], a service of LIGO Laboratory, the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA.

ACKNOWLEDGMENTS

We would like to thank every member of the LVK Collaboration who has taken the time to analyze and discuss the results presented here, as it has helped us understand them better. In particular, we would like to extend special thanks to Colm Talbot, Geraint Pratten, Cecilio García-Quiros, and Soichiro Morisaki for their exceptional comments and feedback. Their input has played a role in shaping and improving the ideas and content of this paper. We would also like to thank Colm Talbot and Jose Antonio Font for their work reviewing this paper within the LIGO and Virgo Collaborations, respectively. They acknowledge support from the research project PID2021-123012NB-C43 and the Spanish Research Agency (Agencia Estatal de Investigación) through the Grant IFT Centro de Excelencia Severo Ochoa No. CEX2020-001007-S, funded by MCIN/AEI/10.13039/501100011033. G.M. acknowledges support from the Ministerio de Universidades through Grant

No. FPU20/02857 and J.F.N.S. acknowledges support from MCIN through Grant No. PRE2020-092571. The authors acknowledge use of the Hydra cluster at the IFT, on which some of the numerical computations for this paper took place. We acknowledge the support of the Supercomputing Wales project, which is partially funded by the European Regional Development Fund (ERDF) via Welsh Government. LIGO Laboratory and Advanced LIGO are funded by the U.S. 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 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. The construction and operation of KAGRA are funded by Ministry of Education, Culture, Sports, Science and Technology (MEXT), 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.

APPENDIX: FAST WAY TO UPDATE

$\|\hat{A}^{-1}\|_F$ AND $\sigma_{\text{EIM}}^{\text{tot}}$

In this appendix, we assume that we have the inverse of the matrix $\bar{A}_{ij} = e_j(X_i)$ and its Frobenius norm $\|\hat{A}^{-1}\|_F$, defined in Eq. (47), and we want to compute the inverse and Frobenius norm of the inverse of the matrix A_{ij} , defined as

$$A_{ij} = \begin{cases} e_j(x_{\beta_i}) & i \neq k \\ e_j(x_{\beta'_k}) & i = k, \end{cases} \quad (\text{A1})$$

which is just the result of changing the row k of \bar{A}_{ij} . We then use the fact that, from the properties of the inverse \bar{A}_{ij} , we have

$$(\hat{A}\hat{A}^{-1})_{ij} = \begin{cases} \delta_{ij} & i \neq k \\ \sum_{l=1}^n e_l(x_{\beta'_k})(\hat{A}^{-1})_{lj} \equiv c_j & i = k. \end{cases} \quad (\text{A2})$$

Since the matrix of Eq. (A2) has such a simple structure, it can be analytically inverted as

$$((\hat{A}\hat{A}^{-1})^{-1})_{ij} = \begin{cases} \delta_{ij} & i \neq k \\ -\frac{c_j}{c_k} & i = k, j \neq k \\ \frac{1}{c_k} & i = j = k, \end{cases} \quad (\text{A3})$$

and we can use that $\hat{A}^{-1} = \hat{A}^{-1}(\hat{A}\hat{A}^{-1})^{-1}$ to show that

$$(\hat{A}^{-1})_{ij} = \begin{cases} (\hat{A}^{-1})_{ij} - \frac{c_j}{c_k}(\hat{A}^{-1})_{ik} & j \neq k \\ \frac{1}{c_k}(\hat{A}^{-1})_{ik} & j = k. \end{cases} \quad (\text{A4})$$

We observe that this way of computing the inverse will require $O(n^2)$ for computing c_j with Eq. (A2) and also $O(n^2)$ operations to update each element of \hat{A}^{-1} using Eq. (A4). So the total number of operations will be $O(n^2)$, much smaller than the $O(n^3)$ required to directly invert the matrix.

Using this expression for the updated inverse, we can find a way to update also the Frobenius norm of the inverse, which is given by

$$\begin{aligned} \|\hat{A}^{-1}\|_F^2 &= \sum_{i,j=1}^n |(\hat{A}^{-1})_{ij}|^2 \\ &= \sum_{i,j=1}^n \left| (\hat{A}^{-1})_{ij} - \frac{c_j}{c_k}(\hat{A}^{-1})_{ik} \right|^2 + \sum_{i=1}^n \left| \frac{1}{c_k}(\hat{A}^{-1})_{ik} \right|^2 \\ &= \|\hat{A}^{-1}\|_F^2 + \frac{1}{|c_k|^2} \left(1 + \sum_{j=1}^n |c_j|^2 \right) \left[\sum_{i=1}^n |(\hat{A}^{-1})_{ik}|^2 \right] \\ &\quad - 2\text{Re} \left(\frac{1}{c_k} \sum_{j=1}^n \left[\sum_{i=1}^n (\hat{A}^{-1})_{ik} (\hat{A}^{-1})_{ij}^* \right] c_j \right), \end{aligned} \quad (\text{A5})$$

where we can precompute with $O(n^2)$ operations the factors in square brackets that only depend on \hat{A} for each row k that we will change, and afterward, updating the Frobenius norm will only need $O(n)$ operations on top of the $O(n^2)$ operations needed to compute c_j for each new row q we want to test. Since with Eq. (A5) we do not need to update the inverse each time that we want to update its Frobenius norm, we can avoid the $O(n^2)$ memory allocations that are needed in Eq. (A4).

We will now also look for a method to rapidly compute $\sigma_{\text{EIM}}^{\text{tot}}$. We assume that we have the value computed for an EIM whose variables we denote with a bar over them,

$$\bar{\sigma}_{\text{EIM}}^{\text{tot}} = \sum_{B=1}^{n_B} (\lambda_B + \|\hat{A}^{-1} \bar{v}_B\|^2), \quad (\text{A6})$$

where we have defined

$$\bar{v}_{B,i} = \sqrt{\lambda_B} u_{B,\beta_i}. \quad (\text{A7})$$

When we change the k th interpolation node of the EIM from β_k to β'_k , this becomes

$$v_{B,i} = \begin{cases} \sqrt{\lambda_B} u_{B,\beta_i} & i \neq k \\ \sqrt{\lambda_B} u_{B,\beta'_k} & i = k, \end{cases} \quad (\text{A8})$$

and the value of multiplying \hat{A} by \vec{v}_B will change to

$$\begin{aligned} (\hat{A}^{-1} \vec{v}_B)_i &= \sum_{j=1}^n (\hat{A}^{-1})_{ij} v_{B,j} \\ &= \sum_{j=1}^n \left(\hat{A}_{ij}^{-1} - \frac{c_j}{c_k} \hat{A}_{ik}^{-1} \right) v_{B,j} + \frac{1}{c_k} \hat{A}_{ik}^{-1} v_{B,k} \\ &= \underbrace{\sum_{j=1}^n \hat{A}_{ij}^{-1} \bar{v}_{B,j}}_{\bar{\Omega}_{B,i}} + \underbrace{\left[\frac{1}{c_k} \left(v_{B,k} - \sum_{i=1}^n c_j \bar{v}_{B,j} \right) \right]}_{\Theta_B} \hat{A}_{ik}^{-1}, \end{aligned} \quad (\text{A9})$$

where we have used the updated value of \hat{A}^{-1} computed in Eq. (A4) and we put a bar over the variables that do not depend on the value of the new interpolation node. Using Eq. (A9), $\bar{\sigma}_{\text{EIM}}^{\text{tot}}$ becomes

$$\begin{aligned} \bar{\sigma}_{\text{EIM}}^{\text{tot}} &= \sum_{B=1}^{n_\lambda} \left(\lambda_B + \sum_{i=1}^n |\bar{\Omega}_{B,i} + \Theta_B \bar{\Gamma}_i|^2 \right) \\ &= \bar{\sigma}_{\text{EIM}}^{\text{tot}} + \left(\sum_{B=1}^{n_\lambda} |\Theta_B|^2 \right) \left[\sum_{i=1}^n |\bar{\Gamma}_i|^2 \right] \\ &\quad + 2\text{Re} \left\{ \sum_{B=1}^{n_\lambda} \left(\Theta_B \left[\sum_{i=1}^n \bar{\Omega}_{B,i}^* \bar{\Gamma}_i \right] \right) \right\}. \end{aligned} \quad (\text{A10})$$

In general, we will assume that $n_\lambda \gg n$. For each row k that we change, we can precompute with $O(nn_\lambda)$ operations all the factors in square brackets that will stay constant. Afterward, the computational complexity of updating the value of $\bar{\sigma}_{\text{EIM}}^{\text{tot}}$ will require $O(nn_\lambda)$ operations for computing Θ_B and only $O(n_\lambda)$ additional operations to evaluate Eq. (A10).

-
- [1] J. Aasi, B. P. Abbott, R. Abbott *et al.* (The LIGO Scientific Collaboration), *Classical Quantum Gravity* **32**, 074001 (2015).
- [2] F. Acernese, M. Agathos *et al.*, *Classical Quantum Gravity* **32**, 024001 (2014).
- [3] T. Akutsu *et al.* (KAGRA Collaboration), *Nat. Astron.* **3**, 35 (2019).
- [4] B. P. Abbott *et al.*, *Living Rev. Relativity* **23**, 3 (2020).
- [5] M. Maggiore, C. V. D. Broeck, N. Bartolo, E. Belgacem, D. Bertacca, M. A. Bizouard, M. Branchesi, S. Clesse, S. Foffa, J. García-Bellido, S. Grimm, J. Harms, T. Hinderer, S. Matarrese, C. Palomba, M. Peloso, A. Ricciardone, and M. Sakellariadou, *J. Cosmol. Astropart. Phys.* **03** (2020) 050.
- [6] M. Evans *et al.*, [arXiv:2109.09882](https://arxiv.org/abs/2109.09882).
- [7] A. Klein, E. Barausse, A. Sesana, A. Petiteau, E. Berti, S. Babak, J. Gair, S. Aoudia, I. Hinder, F. Ohme, and B. Wardell, *Phys. Rev. D* **93**, 024003 (2015).
- [8] N. Tamanini, C. Caprini, E. Barausse, A. Sesana, A. Klein, and A. Petiteau, *J. Cosmol. Astropart. Phys.* **04** (2016) 002.
- [9] S. Babak, J. Gair, A. Sesana, E. Barausse, C. F. Sopuerta, C. P. Berry, E. Berti, P. Amaro-Seoane, A. Petiteau, and A. Klein, *Phys. Rev. D* **95**, 103012 (2017).
- [10] E. Thrane and C. Talbot, *Pub. Astron. Soc. Aust.* **36**, E010 (2019).
- [11] M. Maggiore, *Gravitational Waves. Vol. 1: Theory and Experiments*, Oxford Master Series in Physics (Oxford University Press, New York, 2007), p. 572.
- [12] C. Mills and S. Fairhurst, *Phys. Rev. D* **103**, 024042 (2021).
- [13] S. Morisaki, *Phys. Rev. D* **104**, 044062 (2021).
- [14] B. Zackay, L. Dai, and T. Venumadhav, [arXiv:1806.08792](https://arxiv.org/abs/1806.08792).
- [15] N. J. Cornish, *Phys. Rev. D* **104**, 104054 (2021).
- [16] A. J. K. Chua and M. Vallisneri, *Phys. Rev. Lett.* **124**, 041102 (2020).
- [17] S. R. Green, C. Simpson, and J. Gair, *Phys. Rev. D* **102**, 104057 (2020).
- [18] H. Antil, S. E. Field, F. Herrmann, R. H. Nochetto, and M. Tiglio, *J. Sci. Comput.* **57**, 604 (2013).
- [19] P. Canizares, S. E. Field, J. R. Gair, and M. Tiglio, *Phys. Rev. D* **87**, 124005 (2013).
- [20] R. Smith, S. E. Field, K. Blackburn, C.-J. Haster, M. Pürrer, V. Raymond, and P. Schmidt, *Phys. Rev. D* **94**, 044031 (2016).
- [21] H. Qi and V. Raymond, *Phys. Rev. D* **104**, 063031 (2021).
- [22] S. Morisaki and V. Raymond, *Phys. Rev. D* **102**, 104020 (2020).
- [23] S. Morisaki, R. Smith, L. Tsukada, S. Sachdev, S. Stevenson, C. Talbot, and A. Zimmerman, [arXiv:2307.13380](https://arxiv.org/abs/2307.13380).
- [24] C. Pankow, P. Brady, E. Ochsner, and R. O’Shaughnessy, *Phys. Rev. D* **92**, 023002 (2015).
- [25] J. Lange, R. O’Shaughnessy, and M. Rizzo, [arXiv:1805.10457](https://arxiv.org/abs/1805.10457).
- [26] L. Pathak, A. Reza, and A. S. Sengupta, *Phys. Rev. D* **108**, 064055 (2022).
- [27] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, *Phys. Rev. Lett.* **113**, 151101 (2014).
- [28] G. Pratten, C. García-Quirós, M. Colleoni, A. Ramos-Buades, H. Estellés, M. Mateu-Lucena, R. Jaume, M. Haney, D. Keitel, J. E. Thompson *et al.*, *Phys. Rev. D* **103**, 104056 (2021).
- [29] EigROQ, <https://github.com/gmorras/EigROQ>.
- [30] L. S. Finn, *Phys. Rev. D* **46**, 5236 (1992).

- [31] T. Dietrich, A. Samajdar, S. Khan, N.K. Johnson-McDaniel, R. Dudi, and W. Tichy, *Phys. Rev. D* **100**, 044003 (2019).
- [32] D. Chiaramello and A. Nagar, *Phys. Rev. D* **101**, 101501 (2020).
- [33] P. Jaranowski, A. Krolak, and B. F. Schutz, *Phys. Rev. D* **58**, 063001 (1998).
- [34] M. Pürrer, *Classical Quantum Gravity* **31**, 195010 (2014).
- [35] R. Cotesta, S. Marsat, and M. Pürrer, *Phys. Rev. D* **101**, 124040 (2020).
- [36] D. B. Szyld, *Numerical Algorithms* **42**, 309 (2006).
- [37] C. García-Quirós, S. Husa, M. Mateu-Lucena, and A. Borchers, *Classical Quantum Gravity* **38**, 015006 (2021).
- [38] P. Ajith, M. Hannam, S. Husa, Y. Chen, B. Brügmann, N. Dorband, D. Müller, F. Ohme, D. Pollney, C. Reisswig, L. Santamaría, and J. Seiler, *Phys. Rev. Lett.* **106**, 241101 (2011).
- [39] L. Santamaría, F. Ohme, P. Ajith, B. Brügmann, N. Dorband, M. Hannam, S. Husa, P. Mösta, D. Pollney, C. Reisswig, E. L. Robinson, J. Seiler, and B. Krishnan, *Phys. Rev. D* **82**, 064016 (2010).
- [40] P. Schmidt, F. Ohme, and M. Hannam, *Phys. Rev. D* **91**, 024043 (2015).
- [41] S. S. Wilks, *Ann. Math. Stat.* **9**, 60 (1938).
- [42] S. R. Cook, A. Gelman, and D. B. Rubin, *J. Comput. Graph. Stat.* **15**, 675 (2006).
- [43] S. Talts, M. Betancourt, D. Simpson, A. Vehtari, and A. Gelman, [arXiv:1804.06788](https://arxiv.org/abs/1804.06788).
- [44] I. Romero-Shaw *et al.*, *Mon. Not. R. Astron. Soc.* **499**, 3295 (2020).
- [45] J. S. Speagle, *Mon. Not. R. Astron. Soc.* **493**, 3132 (2020).
- [46] G. Ashton *et al.*, *Astrophys. J. Suppl. Ser.* **241**, 27 (2019).
- [47] G. van Rossum, in *Proceedings of the 2007 USENIX Annual Technical Conference, Santa Clara, CA, USA, 2007*, edited by J. Chase and S. Seshan (USENIX Association, Berkeley, CA, 2007).
- [48] LIGO Scientific Collaboration, LIGO Algorithm Library—LALSuite, free software (GPL) (2018).
- [49] L. Prechelt, *Computer* **33**, 23 (2000).
- [50] B. Abbott *et al.*, *Phys. Rev. Lett.* **119**, 161101 (2017).
- [51] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), [arXiv:2111.03606](https://arxiv.org/abs/2111.03606).
- [52] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J. Lett.* **896**, L44 (2020).
- [53] N. J. Cornish and T. B. Littenberg, *Classical Quantum Gravity* **32**, 135012 (2015).
- [54] T. B. Littenberg and N. J. Cornish, *Phys. Rev. D* **91**, 084034 (2015).
- [55] C. Cahillane, J. Betzwieser, D. A. Brown, E. Goetz, E. D. Hall, K. Izumi, S. Kandhasamy, S. Karki, J. S. Kissel, G. Mendell, R. L. Savage, D. Tuyenbayev, A. Urban, A. Viets, M. Wade, and A. J. Weinstein, *Phys. Rev. D* **96**, 102001 (2017).
- [56] F. Acernese *et al.* (The Virgo Collaboration), *Classical Quantum Gravity* **39**, 045006 (2022).
- [57] L. Sun *et al.*, *Classical Quantum Gravity* **37**, 225008 (2020).
- [58] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), *Astrophys. J. Suppl. Ser.* **267**, 29 (2023).
- [59] J. Lin, *IEEE Trans. Inf. Theory* **37**, 145 (1991).
- [60] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. X* **9**, 011001 (2019).
- [61] LIGO Open Science Center (LOSC) Data release for event GW170817 (2017).
- [62] <https://gwosc.org/>.
- [63] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *SoftwareX* **13**, 100658 (2021).