

Significant challenges for astrophysical inference with next-generation gravitational-wave observatories

A. Makai Baker^{1,2}, Paul D. Lasky^{1,2}, Eric Thrane^{1,2} and Jacob Golomb^{3,4}

¹*School of Physics and Astronomy, Monash University, Clayton, Victoria 3800, Australia*

²*OzGrav: The ARC Centre of Excellence for Gravitational-Wave Discovery, Clayton, Victoria 3800, Australia*

³*Department of Physics, California Institute of Technology, Pasadena, California 91125, USA*

⁴*LIGO Laboratory, California Institute of Technology, Pasadena, California 91125, USA*



(Received 6 March 2025; revised 24 July 2025; accepted 29 September 2025; published 13 November 2025)

The next generation of gravitational-wave observatories will achieve unprecedented strain sensitivities with an expanded observing band. They will detect $\mathcal{O}(10^5)$ binary neutron star (BNS) mergers every year, the loudest of which will be in the band for ≈ 90 minutes with signal-to-noise ratios ≈ 1500 . We show that subtleties arising from the rotation of the Earth and the free-spectral range of gravitational-wave interferometers dramatically increases the complexity of next-gen BNS signals compared to the one-minute signals seen by LIGO-Virgo. Reduced-order quadrature, a compression method currently relied upon to speed up the most expensive BNS calculations, may no longer be effective in determining the astrophysical parameters of next-gen BNS signals. We carry out reduced-order inference on a simulated next-gen BNS signal taking into account the Earth's rotation and the observatories' free-spectral range. We show that reduced-order modeling becomes impractical, and the full problem becomes computationally infeasible, when we include data below ≈ 16 Hz—a part of the observing band that is critical for precise sky localization. We discuss potential paths toward solving this complex problem.

DOI: [10.1103/n6t6-5wn3](https://doi.org/10.1103/n6t6-5wn3)

I. INTRODUCTION

Preparations are underway for the next generation of terrestrial gravitational-wave observatories. The European-based Einstein Telescope [1] is designed to achieve strain sensitivities of $\gtrsim 2 \times 10^{-25} \text{ Hz}^{-1/2}$ with a minimum observing frequency of 1 Hz while the American-led Cosmic Explorer [2] is designed to achieve a similar strain sensitivity to the Einstein Telescope with a minimum frequency of 5 Hz [3]. These next-gen facilities will revolutionize gravitational-wave astronomy. Whereas LIGO-Virgo have so far announced $\mathcal{O}(100)$ gravitational-wave observations [4–6], Cosmic Explorer and the Einstein Telescope will detect most stellar mass binary black hole mergers in the Universe and all but the most distant binary neutron star mergers—about 10^5 every year [2].

Expanding the observing band below the ≈ 20 Hz boundary of current detectors will unlock new science. For example, next-gen observatories will become sensitive to intermediate mass black holes [7], and will observe binary

neutron stars for ≈ 90 minutes compared to the $\lesssim 1$ minute signals observed in LIGO-Virgo. This increased duration can be leveraged to improve the localization of binary neutron stars. During a 90 minute signal, the detectors move a distance of ≈ 1700 km due to the Earth's rotation. This creates a baseline comparable to the ≈ 3000 km distance between LIGO Hanford and LIGO Livingston, which can be used for triangulation with a single detector. Taking advantage of this baseline, a single Cosmic Explorer can localize a binary neutron star to within $\approx 10 \text{ deg}^2$ [8], which is in contrast to the localization of $\gtrsim 8000 \text{ deg}^2$ that has been achieved for events like GW190425, seen by a single LIGO detector [9].¹ Next-gen observatories will also dramatically increase the signal-to-noise ratio (SNR) for binary neutron star signals from $\text{SNR} = 40$ for GW170817 [11] to $\text{SNR} \approx 1600$ for a GW170817-like event as seen by Cosmic Explorer.

In order to measure the parameters of binary neutron stars—such as their sky location, their masses, etc.—gravitational-wave astronomers rely on Bayesian inference; see, e.g., [12]. Bayesian inference is already a serious computational expense for LIGO-Virgo; each event can

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

¹There is no substitute for networks of *two* or more next-gen observatories, which can localize a large fraction of their binary neutron star mergers to $\lesssim 1 \text{ deg}^2$; see, e.g., [10].

take between hours and days to analyze, depending on various factors such as the duration of the signal, the SNR, and the computational cost of the waveform used in the analysis. Relatively longer waveforms require more computations than shorter waveforms, causing inference to take longer. Somewhat counterintuitively, analyzing signals with higher SNRs takes longer because stochastic samplers take longer to converge on the narrow likelihood peak. Reference [13] recently pointed out that the large number of long-duration, high-SNR events observed by next-gen detectors likely require a computing paradigm shift.

Data compression techniques—such as reduced-order quadrature [14], multibanding [15], and relative binning [16,17]—have emerged as essential tools in the battle to control computational costs. The basic idea with all of these methods is to exploit sparsity in the gravitational-wave time series. Multibanding and relative binning use different techniques to coarsely sample the time series, whereas reduced-order quadrature represents the series in terms of a comparatively small number of basis elements. This can sometimes dramatically reduce the number of computations required to estimate the gravitational-wave likelihood function, leading to remarkable speed ups.² For example, LIGO-Virgo regularly use reduced order models to speed up binary neutron star calculations by a factor of ≈ 150 , allowing researchers to obtain results in the span of a few days what could otherwise take a year [14].

In [18], some of us argued that reduced order modeling could be used to manage the immense cost of analyzing a 90-minute-long binary neutron star event like GW170817 [11] observed by a network of next-gen observatories with $\text{SNR} > 1000$. That demonstration was not cheap, requiring 10 hours using 160 high-interconnect cores that were modern in 2021. However, the analysis would have been hopeless without the 10^4 speedup made possible with reduced order methods. Unfortunately, we now believe that demonstration was missing key features.

First, we did not properly take into account effects that arise from the rotation of the Earth over the duration of the signal and the lower free spectral range of next-gen detectors. The free spectral range is the frequency corresponding to the inverse travel time of light in the interferometer arms. The response of the interferometer to gravitational waves with frequencies near the free spectral range is complicated: the amplitude and phase are modulated depending on the direction of the incident waves [19,20]. The effect is largely ignored for LIGO-Virgo, which have a free spectral range frequency of $f_{\text{fsr}} \approx 38$ kHz—safely above the spectral content of astrophysical gravitational waves. However, the long arms of

next-gen observatories such as the 40 km Cosmic Explorer place the free spectral range at ≈ 3750 Hz, which is low enough (it turns out) to significantly affect the response of the detectors to binary neutron star signals. Given that the effect of the free spectral range depends on the direction of the incident gravitational wave, the combination of the free spectral range and rotation of the Earth create a more complicated signal space than we considered in [18].

Second, work by [21] has provided more stringent guidelines for determining if a reduced order model is sufficiently accurate to be used for unbiased inference calculations. Previous work had used the peak of the mismatch between the waveform and reduced-order model evaluated over many parameters to determine whether the reduced-order quadrature likelihood was accurate enough [22]. However, [21] suggests that the accuracy of the reduced-order quadrature rule is determined by the relative error in the log-likelihood ratio.³

In this paper, we seek to test the use of compression methods like reduced order modeling on next-gen gravitational-wave detectors—this time carefully taking into account the rotation of the Earth, the free spectral range, and the strict accuracy requirements from [21]. We show that—in practice—reduced order techniques will not work for binary neutron star signals in next-gen observatories because the size of the reduced order model becomes unwieldy. We hypothesize that data compression techniques are in general unlikely to provide a solution. We discuss various alternate pathways worthy of exploration.

The remainder of this paper is organized as follows. In Sec. II we review the formalism of reduced order methods. In doing so, we aim to demystify for a broad audience what is sometimes regarded as an esoteric topic. In Sec. III, we discuss other forms of compression and compare these techniques to reduced order quadrature. In Sec. IV, we describe our method for determining the computational requirements for binary neutron stars observed by next-gen observatories—taking into account the free spectral range and the rotation of the Earth. (Impatient readers can skip directly to Fig. 1, which shows the GW170817-like-event signal-to-noise ratio achievable with reduced order models—given a fixed memory allotment of 100 Gb.) In Sec. V, we demonstrate Bayesian inference with one of our reduced order models. (We employ a minimum frequency of 26 Hz in order to control the size of the model.) Finally in Sec. VI, we summarize our results and discuss possible solutions.

This paper also includes extensive Appendixes. In Appendix A, we describe how to use the open access

²Another way to understand these compression methods is as follows. The space of time series that correspond to gravitational-wave signals is only a small subset of the space of all time series. Therefore, one can sometimes compress the information in a gravitational-wave time series into a much smaller format.

³To make this relation explicit, we derive an approximate relation between the relative error in the log-likelihood ratio and the maximum signal-to-noise ratio that can be studied with reduced-order quadrature without introducing bias in the posteriors. The relation is given in Sec. IIC and derived in Appendix D.

software ROMBUS,⁴ designed to make it easy for non-specialists to generate reduced order models. In Appendix B, we describe our implementation of time- and frequency-dependent effects needed for analysis of signals in next-gen detectors.⁵

II. FORMALISM

In this section, we describe step by step the construction of a reduced order model for a binary neutron star waveform. We then introduce the reduced order quadrature likelihood and describe techniques for efficient reduced order model construction. We finish with a discussion of the Earth's rotation and the free spectral range and describe why these increase the size of our reduced model. This section is included to provide an introduction for non-experts. Readers interested in the punchline should consider skipping ahead to Sec. IV. The figures referenced in this section are included in Appendix C so as to keep the main body relatively succinct. The ROMBUS Python package is used to construct the reduced order models described in this work. To build these models, we make a number of improvements to ROMBUS as described in Appendix A.

A. Building reduced order models

Our goal is to efficiently calculate a gravitational-wave likelihood that compares strain data d with a template for the observed strain h

$$h(t) = \sum_A F_A(t + \Delta(t)) h_A(t + \Delta(t)). \quad (1)$$

Here, $A = +, \times$ is an index for the gravitational-wave polarization state, h_A is the gravitational wave for polarization state A and F_A is an antenna response factor. The antenna response factors implicitly depend on the location $\hat{\Omega}$ and polarization ψ of the source. For long signals, the antenna response factors vary over the duration of the signal because the detector-frame direction to the source changes as the Earth rotates. The variable Δ is the time delay between the detector and the center of the Earth. For long signals, this time delay implicitly depends on sidereal time and the location of the source due to the rotation of the Earth.

We Fourier transform the product in Eq. (1) to obtain an expression in the frequency domain where we do our likelihood evaluations because the covariance matrix can be approximated as diagonal. Therefore, the frequency domain expression becomes,

$$h(f) = \sum_A \int dt F_A(t) h_A(t) e^{i2\pi f t}, \quad (2)$$

$$= \sum_A \int dt F_A(t) A_A(t) e^{i(2\pi f t - \phi(t))}. \quad (3)$$

Here, we have made use of the post-Newtonian approximation to decompose the strain polarizations into their amplitude $A_{+/\times}(t)$ and phase $\phi(t)$. Then, we apply the stationary phase approximation, which is valid provided the amplitude of the time domain signal varies slowly with respect to the phase. Under this scenario, the Fourier transform can be simplified by expanding the integral around the point of stationary phase t^* , determined by setting the time derivative of the argument of the exponent in Eq. (3) to zero so that,

$$\left. \frac{d\phi(t)}{dt} \right|_{t=t^*} = 2\pi f. \quad (4)$$

Then, the argument of the exponent can be expanded around t^* in a Taylor series to give

$$2\pi f t - \phi(t) \approx 2\pi f t^* - \phi(t^*) - \frac{1}{2} \ddot{\phi}(t^*) (t - t^*)^2 + \dots, \quad (5)$$

where we ignore terms of higher order than $(t - t^*)^2$. Then, the integral in Eq. (3) is [23]

$$h(f) \approx \sum_A F_A(t^*) A_A(t^*) e^{i(2\pi f t^* - \phi(t^*))} \times \int dt e^{-i\frac{1}{2}\ddot{\phi}(t^*)(t-t^*)^2}. \quad (6)$$

Integrating the Taylor expansion with respect to phase, and substituting the strain polarizations for their post-Newtonian expressions, gives the frequency-domain detector response

$$h(f) = e^{i\Psi(f)} \sum_A F_A(f) h_A(f), \quad (7)$$

where $F_A(f) \equiv F_A(t^*(f))$ and we have defined the *phase of the detector response* to be

$$\Psi(f) = 2\pi f t^*(f) - \phi(t^*(f)) - \frac{\pi}{4}. \quad (8)$$

The stationary point t^* can be defined in terms of frequency through the following relation of the orbital frequency and the energy $E(f)$ and flux $F(f)$ of the binary [24]

$$\frac{df}{dt} = -\frac{F(f)}{dE(f)/df}, \quad (9)$$

and is given to lowest order in the post-Newtonian expansion by (setting $G = c = 1$)

⁴<https://github.com/ADACS-Australia/rombus>.

⁵https://git.ligo.org/lscsoft/bilby/-/merge_requests/1370.

$$t^*(f) = t_c - \frac{5}{256} \mathcal{M}_c^{-5/3} (\pi f)^{-8/3}. \quad (10)$$

For a more detailed derivation, we refer the reader to [25]. At low frequencies, incident gravitational wave signals spend hours in the observing band. Over this duration the Earth rotates significantly introducing a frequency dependence in the antenna response factors F_A and phase of the detector response Ψ . At high frequencies, the detector arm length becomes comparable to the wavelength of incident gravitational waves, which also introduces a frequency dependence in the antenna response factors. Now that the detector response in Eq. (7) includes frequency dependent antenna response factors and phase, we can properly account for the effects due to Earth's rotation and the free spectral range.

Reduced order modeling constructs a sparse representation of the gravitational waveform in two steps:

- (1) construct a reduced basis out of a training set of waveforms,
- (2) interpolate the space of waveforms using the reduced basis with minimal loss in accuracy.⁶

Step 1: The reduced basis algorithm selects an N -component, optimal⁷ basis $\{e_i\}$ from a training set of waveforms [27]. Since we are building a reduced order model for gravitational-wave signals from binary neutron stars, each element e_i is an orthonormalized gravitational waveform as measured by an interferometric detector. In order to achieve this goal, the algorithm goes through iterations.

At each iteration, the algorithm projects the training set waveforms onto the reduced basis. The waveform with the worst projection error is added to the basis. This minimizes the *maximum* approximation error of the basis. The algorithm iterates until the maximum approximation error of the basis is less than a user-defined tolerance, which we set to double machine precision $\epsilon = 10^{-16}$. The relationship between maximum approximation error and the number of basis elements is shown in Fig. 6. At $\epsilon = 10^{-14}$, the mismatch—which is defined as a positive-definite quantity—becomes negative due to numerical instabilities and we can no longer reliably compute the error.

The reduced basis method is sometimes preferred over alternative basis construction methods (such as singular value decomposition) because it provides guarantees on the *maximum* approximation error $\epsilon_{\max}$. In the case of gravitational waves, it is important that the error is small

⁶In mathematics, “sparse representation” refers to the idea of representing signals with some small number of “atoms” (component functions).

⁷The reduced basis algorithm is considered optimal in the sense that the decay rate of the approximation error of the basis is bounded by the decay rate of the Kolmogorov N width [26]—the minimum possible approximation error of a manifold V onto an N -dimensional subspace $V_N \subset V$.

compared to the *noise* in a way we quantify below. If this condition is not met, then the approximation error becomes a systematic error that can affect astrophysical inference. Thus, it is important to know $\epsilon_{\max}$ since this determines the maximum signal-to-noise ratio of signals that can be studied without bias. Methods that do not guarantee $\epsilon_{\max}$ “work until they don’t”—that is, they sometimes produce significantly biased results, even if they also sometimes produce reliable results.

The empirical interpolation method uses the reduced basis to construct a sparse representation of the waveform in the frequency domain. The empirical interpolant is referred to as the *reduced order model*. *Reduced order modeling*, meanwhile, refers collectively to the basis construction technique and the use of empirical interpolation.

Step 2: The model can be further compressed by replacing the regularly spaced frequency bins (obtained when one Fourier transforms the gravitational-wave strain time series) with N *frequency nodes* $\{F_i\}$. There are N such nodes—the same as the number of basis elements. This is because the empirical interpolation algorithm selects a unique node for each basis element to minimize the maximum approximation error of the interpolant. The interpolant is sparse when N is less than L , the number of data points in the original time series,

$$L = T_m(f_{\max} - f_{\min}). \quad (11)$$

Here, T_m is the duration of the signal/data—which is typically rounded to the next-highest power of two—and $(f_{\min}, f_{\max})$ are the minimum and maximum frequencies. In practice, reduced order models in gravitational-wave astronomy are extremely sparse so that $N \ll L$.

The empirical interpolant is [28]

$$h^{\text{interp}}(f, \theta) = \sum_{j=1}^N h(F_j, \theta) B_j(f), \quad (12)$$

where B_j are elements of an interpolation matrix constructed from the reduced basis elements.⁸ Here θ are the binary parameters like component masses, distance to the source, etc. The elements of the interpolation matrix are defined as

$$B_j(f) = \sum_i e_i(f) (V^{-1})_{ij}, \quad (13)$$

where $\{e_i\}$ are the orthonormalized reduced basis elements and V is a matrix with elements defined by

⁸Interestingly, the interpolation matrix and frequency nodes in Eq. (12) can be used to construct reduced order model representations for other models that are sufficiently similar to the original model, i.e., other models that are in the span of the reduced basis.

$$V_{ij} = e_i(F_j). \quad (14)$$

The accuracy of the empirical interpolant is typically 100 times worse than the maximum reduced basis error for gravitational waveforms [29]. Reduced order models can be used, as in this work, to form a compressed likelihood. However, reduced order modeling can also be used to build *surrogates* for fast waveform evaluation, enabling the use of otherwise prohibitively slow waveforms (see, e.g. [30]).

B. Reduced order quadrature

The standard log-likelihood used for gravitational-wave parameter estimation is

$$\ln \mathcal{L} = \langle d, h \rangle - \frac{1}{2} \langle h, h \rangle - \frac{1}{2} \mathcal{Z}_n, \quad (15)$$

where d is the data, h is a template waveform, and $\mathcal{Z}_n \equiv \langle d, d \rangle$ is the noise evidence. Following the notation from [12], $\langle a, b \rangle$ is the noise-weighted inner product of a with b , which includes a sum over evenly spaced frequency bins,

$$\langle a, b \rangle \equiv \frac{4}{T} \sum_{j=1}^L \frac{\text{Re}(a_j^* b_j)}{S_h(f_j)}. \quad (16)$$

Here, $S_h(f)$ is the single-sided strain noise power spectral density.

Evaluating the sum in Eq. (16) is computationally expensive. By replacing the waveform h with the reduced order model in Eq. (12), one can instead sum over sparse elements. The resulting expression is referred to as the *reduced order quadrature* (ROQ) likelihood,

$$\ln \mathcal{L} = \sum_{j=1}^{N_L} h(F_{L,j}, \theta) \omega_j - \frac{1}{2} \sum_{i=1}^{N_Q} |h(F_{Q,i}, \theta)|^2 \psi_i - \frac{1}{2} \mathcal{Z}_n. \quad (17)$$

The quantities ω_j and ψ_i are integration weights that can be computed in preprocessing. The variables N_L and N_Q are the number of basis elements required for the reduced order model of the *Linear* and *Quadratic* detector response terms in Eq. (17). F_L and F_Q are the frequency nodes belonging to the linear response term and the quadratic response term, respectively.

In this work we build a reduced order model that can be compared directly with the data as in Eq. (1). This is in contrast to previous work by [18,21,22], which built models for $h(t)$ ignoring the effects due to the free spectral range and only including the change in detector orientation due to Earth's rotation. These studies ignore the time dependent time delay $\Delta(t)$ that arises from the change in the light travel time from the Earth's geocenter to the detector location as Earth rotates. This allows the phase to be absorbed into the weights resulting in smaller reduced order models (ROMs), but requires building the weights for $\mathcal{O}(10^5)$ parameter

values to interpolate between. Instead, we incorporate the phase directly into the empirical interpolant. This is necessary to include the changing light travel time and therefore fully include the effects of Earth's rotation. The phase becomes time dependent and rapidly varying, and can no longer be straightforwardly absorbed into the weights. This means we no longer need to interpolate the weights which greatly reduces the startup cost of reduced order inference.

For a typical binary neutron star signal in Cosmic Explorer with $\mathcal{M}_c = 1.385 M_\odot \pm 1 \times 10^{-4} M_\odot$ starting from $f_{\min} = 5$ Hz and ignoring Earth's rotation and the free spectral range effects, only a few hundred basis elements are required and the likelihood speedup using ROQ is

$$\frac{L}{N_L + N_Q} \sim \mathcal{O}(10^4). \quad (18)$$

For details on the derivation of the reduced order quadrature likelihood in Eq. (17) and how it differs to those used in other works, we refer the reader to Appendix D 2.

C. Accuracy of ROQ

Having covered how to construct a reduced order likelihood, we now discuss how to determine if the reduced order approximations are adequate for inference calculations at a given signal-to-noise ratio. To do so, we employ an approximate relation between the relative error in the log-likelihood ratio and the maximum signal-to-noise ratio ρ able to be studied by a reduced order quadrature rule,

$$\frac{\Delta \ln \mathcal{L}}{\ln \mathcal{L}} \lesssim \frac{1}{\rho^2}. \quad (19)$$

The likelihood error is defined as

$$\Delta \ln \mathcal{L} = \ln \mathcal{L} - \ln \mathcal{L}_{\text{ROQ}}, \quad (20)$$

where $\ln \mathcal{L}$ is the standard likelihood and $\ln \mathcal{L}_{\text{ROQ}}$ is the ROQ likelihood. This rule is demonstrated in Appendix D.

D. Constructing the training set

Our model is constructed using the IMRPhenomPv2_NRTidalv2 approximant. Millions of waveform evaluations are necessary to construct an accurate model. We utilize a number of strategies in order to control the size of the training set: (1) adaptively sample the frequency domain, (2) iteratively enrich the training set, and (3) utilize targeted sampling techniques and a noise-weighted inner product to build the reduced basis.

Adaptive sampling. The first strategy we use to minimize the size in memory of the training set is evaluating waveforms over adaptively sampled frequency domains. The steps are outlined below:

- (i) Split the frequency domain into bands b with maximum frequencies according to 2^n : 5 Hz, 8 Hz, 16 Hz, etc. Since the inspiral signal is quasimonochromatic, it is a good approximation to say that the signal is in each band for only a fraction of the total signal duration. This allows us to downsample each band according to the amount of time that the signal is present in each band. Downsampling reduces the number of frequency bins, which makes the calculation more tractable.
- (ii) Downsample each band with an adaptive resolution $\Delta f^{(b)}$ based on the time spent in each band by the lowest mass binary in parameter space. Thus, for example, the 5–8 Hz band has a resolution of $\Delta f = 1.2 \times 10^{-4}$ while the 32–64 Hz band has a resolution of $\Delta f = 1.5 \times 10^{-2}$. The 5–8 Hz band requires higher resolution than the 32–64 Hz band because the binary spends 128 times longer in the former than the latter.
- (iii) Build the reduced basis from a training set of downsampled waveforms. In this work, we utilize a training set sizes of $\sim 10^6$ for reduced order models starting from $f_{\min} \geq 16$ Hz and $\sim 2 \times 10^5$ for the reduced order model starting from $f_{\min} < 16$ Hz. This is due to computational limitations discussed in Sec. II E.
- (iv) Take the (few) parameters $\{\theta_i\}_i^{N_L}$ that give the downsampled basis elements. Build a training set of waveforms evaluated over the full frequency array L at these parameters. We now have a faithful training set. As this training set is only a few hundred faithful waveforms, it is relatively inexpensive compared to directly constructing a faithful training set of millions of waveforms.
- (v) Build the faithful reduced basis over this new training set. Because the downsampled waveforms were sampled above the Nyquist rate, we can safely assume the parameters identified by the downsampled reduced basis accurately represent the most informative waveforms in the faithful case. We validate this by computing the mismatch between the reduced order model and the true waveform for 50,000 parameter values.

This adaptive sampling technique reduces the memory requirements of the training set by $\mathcal{O}(10^2)$ and enables fast construction of faithful bases.

The time in each band is determined using the post-Newtonian expansion of the time to merger taken to second order, which is an adequate approximation for the early inspiral,

$$T_m = \frac{5}{256} \mathcal{M}_c^{-5/3} (\pi f)^{-8/3} \propto f^{-8/3}, \quad (21)$$

where $\mathcal{M}_c$ is the chirp mass. The post-Newtonian approximation breaks down near the innermost stable circular orbit. Therefore, we limit the maximum frequency resolution to $\Delta f^{(b)} = 1$ to ensure that we do not under-sample the high-frequency bands. To avoid sampling below the Nyquist rate, the time spent in each band is rounded to the nearest power of two. When building the reduced basis over the faithful training set, we track the projection errors to validate the adaptive sampling procedure.

Iterative enrichment. The second technique we employ to minimize the size of the training set is iterative training set enrichment. The basic idea is to validate our model over multiple training sets. It is unlikely that two randomly constructed training sets will have the same waveforms, so this effectively increases the number of training set waveforms we build our model over. We build three supplementary training sets, which is enough to ensure model accuracy. Each training set is projected onto the reduced basis, and any waveform with approximation errors greater than the user-defined tolerance ϵ is added to the basis.

Targeted sampling. Finally, we draw waveforms from regions of parameter space that we know are important for accurate models. We draw half of our waveforms randomly from the standard gravitational-wave prior.⁹

The standard BILBY ROQ likelihood assumes t_c is fixed during ROM construction and precomputes ROQ weights over a grid in time before parameter estimation. The time dependence of the detector response in Eq. (7), and therefore the ROQ likelihood in Eq. (15), is less straightforward after including the effects due to the Earth's rotation and free spectral range (FSR), in part because the detector response phasing depends on the function $\Psi(f)$. So, reintroducing the time dependence during ROQ weight computation is more difficult. Therefore, we choose to sample over time in the ROM construction in order to provide a more straightforward implementation of the ROQ likelihood Eq. (15).

The luminosity distance d_L is typically fixed during ROM construction as the training set waveforms are normalized. If the luminosity distance does not affect the overall structure of the waveform, it should not affect basis size. So, we sample over luminosity distance for peace of mind (granted, the added dimensions may make the training set construction more difficult).

Following [22], the other half of our waveforms are drawn from a regularly spaced grid. The grid is uniform in all parameters other than chirp mass, which we sample uniformly in $\mathcal{M}_c^{-5/3}$ as this term describes the inspiral

⁹We set uniform priors on all parameters. The chirp mass prior is given by $1.18 \leq \mathcal{M}_c \leq 1.18 + 5\Delta\mathcal{M}_c$, where $\Delta\mathcal{M}_c = 8 \times 10^{-3} \times (32/\text{SNR})$. We limit the mass ratio to $0.5 \leq q \leq 1$, the dimensionless spin magnitudes to $|a_{1/2}| \leq 0.05$, the luminosity distance to $1 \leq d_L \leq 4000$ Mpc, the tidal deformabilities to $0 \leq \Lambda_{1/2} \leq 1000$, and the time of coalescence to $|t_c| \leq 0.1$ s. We use the full range for all other parameters.

phase at leading order. There are 500 grid points in each dimension. We draw randomly from the multidimensional grid. Drawing points from a grid ensures that half of the training set waveforms are uniformly spaced, which ensures a more even sampling of the parameter space.

It has also been shown that basis elements are preferentially selected from the boundary of parameter space [18]. This may arise from the slowly varying and smooth nature of the space of waveforms, such that the most distinct waveforms lie on opposite ends of the parameter space boundary. Therefore, all boundary parameter samples are included in the training set. We also construct the reduced basis using the noise-weighted inner product. Then, the greedy algorithm selects basis elements that are more informative with respect to the likelihood, further decreasing the size of the basis.

E. Earth's rotation and the free spectral range

Over the hours-long signals expected to be observed by next-gen detectors, the detector moves through space as Earth rotates. Furthermore, as the arm length of the detector L_{arm} approaches the wavelength of the incident wave, the signal evolves over the travel time of photons in the detector arms. The frequency at which this occurs is the FSR, given by

$$f_{\text{FSR}} = \frac{c}{2L_{\text{arm}}}. \quad (22)$$

Here, c is the speed of light. These two phenomena—the rotation of the Earth and the FSR—make the antenna response functions $F^{+/\times}$ and the detector response phase Ψ [Eq. (8)] frequency dependent. The antenna response is given by [20]

$$F^{+,\times}(\hat{n}|t) = D_{ij}(\hat{n}|t)\epsilon_{+,\times}^{ij}(\hat{n}), \quad (23)$$

where D_{ij} is the detector tensor, $\epsilon_{+,\times}^{ij}$ is the polarization tensor, $\hat{n}$ is the direction to the gravitational-wave source, and t is time.¹⁰ In the frequency domain, Eq. (23) becomes,

$$F^{+,\times}(\hat{n}|f) = D_{ij}(\hat{n}|f)\epsilon_{+,\times}^{ij}(\hat{n}). \quad (24)$$

Incorporating the effects due to Earth's rotation and the free spectral range causes both the magnitude and phase of $D_{ij}(f)$ to vary with frequency. This in turn means that the amplitude and phase of $F^{+,\times}(\hat{n}|f)$ varies with frequency. This is in contrast to analyses carried out on LIGO/Virgo data where one can safely approximate $F^{+,\times}(\hat{n}|f)$ as a constant in frequency. This frequency dependence is illustrated in Fig. 5 of Appendix C, where we compare

¹⁰The indices i and j run over the three spatial dimensions while $+$ and $\times$ describe the two polarization states of gravitational waves. Here, the repeated indices imply summation.

the frequency-dependent and static antenna response and phase.

The reduced basis algorithm selects basis elements based on unique features. The varying phasing of the detector response introduces complexity in the structure of the waveforms. For one thing, the frequency dependence is different for different directions $\hat{n}$. The rapidly varying structure of the waveforms causes the basis to become extremely large, making it impractical with current computational resources to construct accurate ROMs over full parameter domains for $f_{\text{min}} \leq 17$ Hz. As a result, the reduced order modeling techniques adopted in the literature currently cannot provide fast and accurate inference for next-generation gravitational wave detectors. We demonstrate this in detail in Sec. IV, where we construct reduced order models across all 17 parameters that parametrize a binary neutron star waveform.

III. ALTERNATIVES TO REDUCED ORDER MODELING

Before we demonstrate the limitations of reduced order modeling, we pause to comment on other methods that have been proposed to control the computational cost of inference. The most notable alternatives to reduced order modeling are multibanding [15] and relative binning [17,31]. Perhaps relative binning and multibanding could be used to enable inference in next-generation detectors, however we are worried that these techniques cannot achieve better compression than reduced order modeling without sacrificing accuracy. Qualitatively, our argument can be understood as follows. Reduced order methods, multibanding, and relative binning are all methods of data compression. However, the complexity arising from the FSR and the rotation of the Earth appears to be too much for reduced order modeling to overcome. There is a finite amount of compression possible based on the variability of waveforms in our signal space. Reduced order modeling appears to provide the most efficient compression scheme, with a guaranteed minimum error. We hypothesize that other compression schemes cannot overcome this added complexity without sacrificing precision. As always, there is no free lunch.

In addition, reduced order quadrature is preferable as it provides *a priori* guarantees on the maximum signal-to-noise ratio that can be studied without introducing bias in the inference.

IV. REDUCED ORDER MODELING WITH LIMITED MEMORY

In this section we show that reduced order modeling—using currently available computers—is not a practical solution for the loudest binary neutron star events observed with next-gen observatories. It is, in fact, extremely challenging to analyze a single event like GW170817

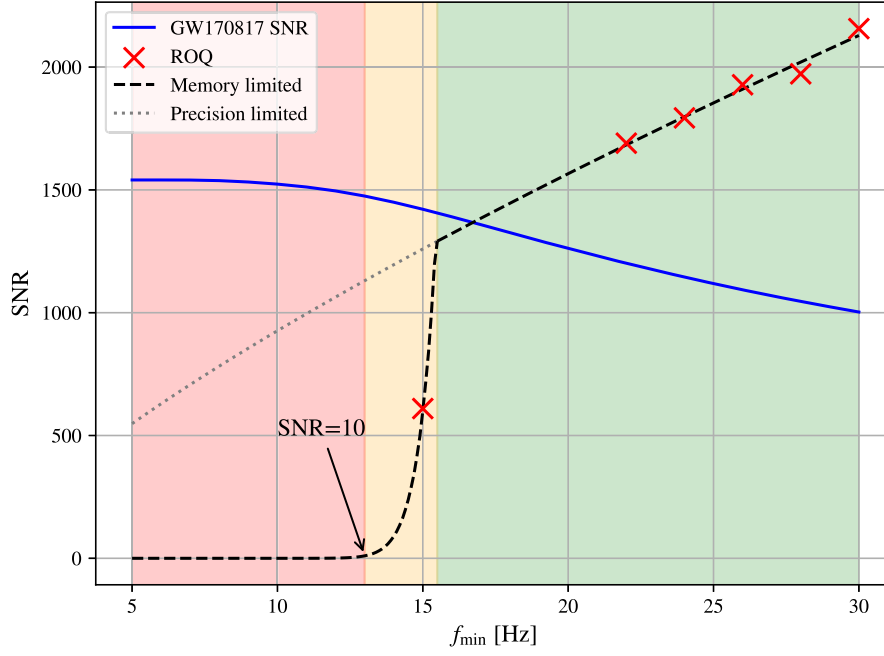


FIG. 1. Limitations of inference with data compression strategies. Solid blue shows the SNR of a GW170817-like event as observed by a single Cosmic Explorer. This is contrasted with the dashed black showing the maximum signal-to-noise ratio that we can currently achieve with reduced order modeling. In the green region above 16 Hz, the maximum SNR is limited only by double machine precision, which is almost sufficient. In the orange region between 13 and 16 Hz, the maximum SNR is limited by our memory limit of 100 Gb; we can carry out inference, but only by artificially inflating error bars. In the red region below 13 Hz, the maximum SNR drops below 10, and we are unable to carry out meaningful calculations, even by inflating error bars.

without throwing away data below ≈ 16 Hz in order to make the calculation tractable.

We construct a set of reduced order models for a binary neutron star signal with different minimum frequencies $f_{\min}$. We require that each model is no more than 100 Gb in size, and that it can be stored using variables of double precision. Anything larger than 100 Gb becomes unwieldy. By limiting the size of our models to 100 Gb, we limit the number of basis functions N , which in turn limits the accuracy of the model. The accuracy of the model, meanwhile, determines the maximum SNR that we can analyze without obtaining biased results via Eq. (19). Thus, the 100 Gb imposes an indirect limit on the maximum SNR.

As we decrease $f_{\min}$, the signals spend more time in band; the time to merge scales according to Eq. (21). Moreover, as we decrease $f_{\min}$ the signal morphology becomes more complicated due to the rotation of the Earth. Thus, each basis element becomes larger and more basis elements are required to achieve a fixed accuracy. As a result, the maximum SNR decreases with $f_{\min}$. Our goal is to determine the relationship between maximum SNR and $f_{\min}$. We compare the maximum SNR with the expected SNR of a GW170817-like binary neutron star with masses $m_{1,2} = 1.4M_{\odot}$ at a distance of 40 Mpc. For a reduced order model to work, we need the maximum SNR to exceed the SNR from a GW170817-like event. We study minimum frequencies between 5 and 30 Hz. Over this range, the

expected SNR of a GW170817-like event varies from $\text{SNR} \approx 1600$ at 5 Hz to $\text{SNR} \approx 1000$ at 30 Hz; see the solid blue curve in Fig. 1.

The first step is to determine the range of chirp mass ΔM_c over which the reduced order model should be built for a given value of $f_{\min}$. We want ΔM_c to be as small as possible in order to minimize the size of the model. However, it must be large enough to safely include the tails of the posterior distribution. We choose ΔM_c to include ± 5 standard deviations from the true value of M_c .¹¹ The size of one standard deviation depends primarily on the signal-to-noise ratio [23], so we estimate ΔM_c with the scaling relation,

$$\Delta M_c \propto \frac{1}{\text{SNR}}. \quad (25)$$

We know $\Delta M_c = 8.3 \times 10^{-3} M_{\odot}$ for GW170817, which was detected with $\text{SNR} = 32$; we use this data point to estimate ΔM_c for other SNR values.

The next step is to build the reduced order model with sufficient accuracy to analyze the GW170817-like event at

¹¹Of course, in an actual analysis, the true value of M_c is unknown, but a maximum-likelihood estimate is provided from a matched filter search upstream of the parameter estimation analysis.

the expected SNR given by the blue curve in Fig. 1. We add basis functions until either (1) we succeed and the accuracy is limited by double machine precision or (2) we fail because the model exceeds 100 Gb.

For $f_{\min} > 16$ Hz, we are interested in the behavior of the accuracy of our reduced order models when limited only by numerical precision. So, we build our reduced order models starting with a high proposal SNR of 2400 Hz as utilized in [22]. The step to infer the maximum SNR for $f_{\min} > 16$ Hz is as follows:

- (1) Set the target SNR $\text{SNR}_{\text{target}}$ to 2400 Hz and build the ROM to be as accurate as possible (an error tolerance of $\epsilon = 10^{-16}$).
- (2) Compute the relative difference $\Delta \ln \mathcal{L} / \mathcal{L}$ in the reduced order quadrature likelihood for 50,000 random waveforms drawn from the prior.
- (3) Using Eq. (19), determine the maximum SNR SNR_{max} from the peak of the $\Delta \ln \mathcal{L} / \mathcal{L}$ distribution.
- (4) If SNR_{max} does not align with the initial $\text{SNR}_{\text{target}}$, repeat from Step 1 with the target SNR set to the SNR_{max} as determined from Step 3.

This process repeats until SNR_{max} from Step 3 matches the initial guess, $\text{SNR}_{\text{target}}$. The procedure is a success and we construct models limited by machine precision, which are less than 100 Gb in size. The relationship between maximum SNR and $f_{\min}$ above 16 Hz is marked on Fig. 1: red crosses show reduced order models we have successfully constructed while the black dashed line shows the empirical scaling relation,

$$\text{SNR}_{\text{max}} \propto f_{\min}^{0.71}. \quad (26)$$

This scaling relation will prove useful momentarily as we seek to extrapolate to lower values of $f_{\min}$ where it is more computationally challenging to construct a reduced order model. In these regions, we use our scaling relations to provide the initial guess $\text{SNR}_{\text{target}}$ for constructing our reduced order models.

Between 16 and 17 Hz, the procedure succeeds in staying under the memory requirement, but the model falls just short of the required accuracy due to the machine precision of our double arrays. This causes the dashed maximum SNR curve in Fig. 1 to fall slightly below the solid blue curve showing the SNR of a GW170817-like event. However, this machine precision limitation is a relatively minor problem compared to the problems that arise below 16 Hz. Below this $f_{\min}$, the analysis becomes limited by memory.

Below 16 Hz, we reduce the SNR by artificially inflating the detector noise. However, reducing the SNR necessitates that we broaden the chirp mass prior according to Eq. (25). This leads to a chicken-and-egg problem: we need the maximum SNR to know how wide to make ΔM_c so we can

calculate the maximum SNR. Fortunately, we can use empirical scaling laws to determine ΔM_c and SNR.

The first scaling relation that we need is one between the number of basis elements B_{max} —required for a model rated up to SNR_{max} —and the minimum frequency:

$$B_{\text{max}} \propto f_{\min}^{-2.2}. \quad (27)$$

Together, Eqs. (26) and (27) imply that

$$\text{SNR}_{\text{max}} \propto B_{\text{max}}^{-0.32}. \quad (28)$$

However, note: this is only true when we hold fixed the error tolerance $\epsilon = 10^{-16}$ as required by machine precision.

Next we need an empirical relationship between SNR and B for *variable* error tolerance. We construct models for three values of $f_{\min}$: 26 Hz, 28 Hz, and 30 Hz. For each $f_{\min}$, we construct models for $\text{SNR} = \{250, 500, 750, 1000\}$ using variable error tolerances ϵ (and thus variable chirp mass ranges ΔM_c). For each of these $f_{\min}$ we fit the relation between SNR and basis size assuming a power law. We obtain the following relation:

$$\text{SNR} = \kappa(f_{\min}) B^{8.3-2.0(f_{\min}/10 \text{ Hz})}. \quad (29)$$

Here, κ depends on $f_{\min}$, but—since we are about to apply this equation at a fixed $f_{\min}$ — κ is a constant for our purposes.

The final thing we need to do before we can use these scaling relations to calculate the relationship between SNR and $f_{\min}$ is to relate the number of basis elements B —and the size of each element—to the *size* of the basis—which we require to be less than 100 Gb. The size of the basis in bytes is

$$S = 16 B T_m (f_{\text{max}} - f_{\min}). \quad (30)$$

Here $f_{\text{max}} - f_{\min}$ is the bandwidth (we take $f_{\text{max}} = 2048$ Hz), 16 is the number of bytes in each double-precision variable used to represent the waveform, and the time to merge is given in Eq. (21). Our requirement that $S < 100$ Gb implies a maximum value for B given by

$$B_{100} = \frac{10^{11}}{16 T_m (f_{\text{max}} - f_{\min})}. \quad (31)$$

Finally, we are ready to calculate the memory-limited signal-to-noise ratio, denoted SNR_{100} , as a function of minimum frequency for $f_{\min} < 16$ Hz—the region where the maximum SNR model exceeds 100 Gb. The steps are as follows:

- (1) Use the scaling relation in Eq. (29) to relate $(B_{100}, \text{SNR}_{100})$ to $(B_{\max}, \text{SNR}_{\max})$. This yields,

$$\text{SNR}_{100} = \left(\frac{B_{100}}{B_{\max}} \right)^{8.3-2.0(f_{\min}/10 \text{ Hz})} \text{SNR}_{\max}. \quad (32)$$

- (2) Determine B_{100} : the number of basis elements required for a 100 Gb basis using Eq. (31).
- (3) Determine $\text{SNR}_{\max}$: the maximum signal-to-noise ratio allowed by machine precision using Eq. (26).
- (4) Determine $B_{\max}$: the number of basis elements associated with $\text{SNR}_{\max}$ using Eq. (27).

Putting everything together yields the following scaling relation for $f_{\min} < 16$ Hz:

$$\text{SNR}_{100} \propto (f_{\min})^{40-9.8(f_{\min}/10 \text{ Hz})} \text{SNR}_{\max}(f_{\min}) \quad (33)$$

This scaling leads to the precipitous drop in the dashed black curve below 16 Hz in Fig. 1. The black dashed curve can be contrasted with the dotted grey curve, which shows the $\text{SNR}_{\max}$ versus $f_{\min}$ if we had no 100 Gb limit so that the models are limited only by machine precision.

To validate these calculations, we build a ROM for $f_{\min} = 15$ Hz where we set our initial SNR guess to be $\text{SNR}_{\text{target}} = 601$, inferred from our scaling relations. We determine the maximum SNR to be $\text{SNR} = 608$ which matches the theoretical prediction of $\text{SNR} = 601$. The size of the basis is 33 Gb in memory, which is relatively close to our 100 Gb limit. Therefore our scaling relations predict the memory to within a factor of a few, which is adequate for the purposes of demonstrating a problem that is many orders of magnitude larger. This 15 Hz ROM is indicated by a red cross in Fig. 1.

We divide Fig. 1 into three colored regions. In the green region above $f_{\min} = 16$ Hz, we can successfully build models with sufficient accuracy to analyze GW170817-like events. In the yellow region between $f_{\min} = 13$ and 16 Hz, we can construct models with $\text{SNR} > 10$ by artificially inflating the noise. This is undesirable because we are throwing away information, but it shows that the calculation is at least close to tractable. In the red $f_{\min} < 13$ Hz region, the SNR falls below 10, and it becomes difficult to construct useful models. A network SNR of 12 is typically required for an event to be considered a detection. Even for a single event, it is therefore challenging to carry out inference down to these $f_{\min}$ values using contemporary computers and with existing data compression strategies.

V. INFERENCE ON A BINARY NEUTRON STAR SIGNAL

In this section we demonstrate inference on a binary neutron star signal like GW170817, taking into account the rotation of the Earth and the FSR. Previous efforts to accurately perform inference on a binary neutron star signal observed by a next-generation detector missed key details.

Reference [22] ignores the effects due to the free spectral range, which contributes to the prohibitive computational expense in construction ROMs. To our knowledge, this is the first demonstration of inference on a binary neutron star signal in next-generation detectors done using a reduced order model—in order to guarantee the likelihood is sufficiently accurate for reliable inference—and including the effects of the rotation of the Earth and the FSR.

We use the reduced order model starting from $f_{\min} = 26$ Hz (safely in the Fig. 1 green zone) with a maximum frequency of 2048 Hz and a signal duration of 128 s. The linear and quadratic reduced bases have $N_L = 2667$ and $N_Q = 2425$ elements, respectively, giving a theoretical likelihood acceleration of 50 (Eq. (18)). We inject a binary neutron star signal generated using the IMRPhenomPv2_NRTidalv2 approximant [32] at $d_L = 40$ Mpc into a single Cosmic Explorer located at the site of the current LIGO Hanford observatory with a zero-noise realization of Gaussian noise. The injection step includes effects from the rotation of the Earth and the FSR. This yields $\text{SNR} = 792$, which is well within the accuracy bounds of the ROQ rule.

We assume standard priors on the 17 parameters that characterize a binary neutron star merger. We utilize uniform priors on chirp mass $\mathcal{M}_c$, mass ratio q , spin magnitudes $\chi_{1,2}$, and the tidal deformabilities $\Lambda_{1,2}$. We adopt an isotropic prior for the direction of the spin vectors. Furthermore, we assume a prior that is uniform in comoving volume and source frame time for the luminosity distance d_L . We adopt standard priors for the other extrinsic parameters. By injecting into zero noise, we expect the posterior distribution to peak at the injected parameter values. This serves as a useful diagnostic to verify the accuracy of the ROQ likelihood in Eq. (17) and the ROMs constructed in this work.

We use the DYNesty [33] nested sampling package as implemented within BILBY [34,35] to estimate the posterior distribution. We use 2000 live points with the acceptance walk Markov chain Monte Carlo method from [36]. The sampling is parallelized over 124 processes. The inference takes 2.3 hours (285 CPU hours) during which the log-likelihood is called 50M times. The average log-likelihood evaluation time is 0.25 s. Therefore, the total sampling time without using reduced order quadrature would be approximately 3527 CPU hours, or 28.4 hours using 124 processes. The reduced order quadrature likelihood accelerates inference from 3527 CPU hours to 285 CPU hours, giving an empirical inference speedup of a factor of 12—less than the theoretical prediction of 50 due to overheads in the inference procedure.

In Figs. 2–4, we show in blue the one and two dimensional posterior distributions for various combinations of parameters. The posterior distributions peak at the true parameter values, indicated in orange, validating the reduced order inference procedure outlined in this work.

The location of the source in the sky is constrained to 403 deg^2 at the 90% credibility. This sky area is far larger than the field of view of the upcoming Vera Rubin observatory, planned to be used for electromagnetic follow-up observations of binary neutron star mergers [37]. Narrow constraints on sky localization are crucial for rapid multimessenger observations. Whilst we expect that the constraints on sky location will improve significantly when considering signals beginning from $f_{\min} = 5 \text{ Hz}$, we face severe limitations in achieving fast and accurate inference below $f_{\min} = 17 \text{ Hz}$.

VI. CONCLUSION

Next-gen gravitational wave detectors will observe thousands of binary neutron star signals, some of which will be in the detecting band for $\gtrsim 1 \text{ hr}$. However, after taking into account subtle effects from the rotation of the Earth and the free spectral range, we show that it is challenging to analyze even one of these long signals with reduced order modeling, and we are concerned that other data analysis strategies may encounter similar challenges to achieve the accuracies required by next-gen detectors.

Our findings exacerbate computational issues pointed out in [13], who estimate the number of cores required to process the hundreds of thousands of compact binary signals detected by next-gen observatories. Even using data compression techniques like reduced order modeling, but assuming no paradigm-changing new algorithms, [13] find that 10^9 – 10^{15} contemporary CPU cores are likely required to analyze one month of data. While [13] focus on the sheer volume of binary signals—and we agree with their assessment that the large volume of detections is problematic—here we show that it is currently challenging to use reduced order modeling to analyze even just *one* of the loudest binary neutron stars that will be detected by Cosmic Explorer and the Einstein Telescope.

In addition to the raw computational challenges posed by next-gen observatories, the community faces formidable challenges relating to systematic error/model mis-specification [38,39]. Key systematic errors include systematic errors in gravitational waveforms, imperfections in the noise model, and calibration error. While the past decade has seen great strides in the development of fast gravitational waveform approximants, the current level of systematic error—which we can estimate by looking at the differences between different approximants—is large compared to the statistical uncertainty with which the loudest binary signals will be observed by next-gen observatories [40].

Likewise, our ability to accurately describe detector noise is limited by the non-Gaussian character of real interferometer noise (see, e.g., [41]). And even Gaussian noise models are complicated by subtle effects, which are often ignored in LIGO/Virgo analyses, but which will be significant systematic errors for next-gen observatories [42–48].

Researchers have proposed elegant solutions to marginalize over uncertainty in the noise model and to take into account edge effects, but these solutions tend to increase the computational cost of the analysis, exacerbating the problems identified here.

Finally, the calibration uncertainty is expected to be a significant source of systematic error for next-gen observatories [49]. Our point is that—even if we are able to solve the computational problems that make it a challenge to carry out inference on long gravitational-wave signals—it may be necessary to significantly inflate error bars due to systematic errors.

We end on an optimistic note. First, some of the challenges associated with next-gen observatories seem to be under control. While LIGO/Virgo detect gravitational-wave signals that are clearly separated in time, Cosmic Explorer and the Einstein Telescope will typically observe multiple binary neutron stars at the same time [50]. However, results by [51] indicate that the presence of neighboring signals is unlikely to interfere with astrophysical inference. And work by [52] suggests it is possible to measure a primordial gravitational-wave background obscured by a foreground of compact binary signals.

Second, while we are pessimistic about the possibility of carrying out inference on long gravitational-wave signals with the current paradigm, we think it may be possible with the development of new techniques. We argue here that the calculations have reached the maximum benefit possible by compression. However, significant speedups may be possible through deep learning schemes (e.g., [53]), which perform amortized inference without the need for likelihood evaluations. In fact, recent work has demonstrated the capabilities of deep learning methods to provide real time inference for binary neutron star signals in current detector networks [54]. Additionally, a machine learning based workflow has been suggested for the efficient analysis of long-duration binary neutron star signals in next-generation detectors [55]. However, the methods outlined in [54,55] utilize multibanding and relative binning to accelerate training their workflows on binary neutron star signals. The performance of these methods is then dependent on the accuracy of these underlying compression techniques. Finally, clever parallelization of the analysis may make it possible to work with manageable-sized reduced order models. We leave this for future work.

ACKNOWLEDGMENTS

This material is based upon work supported by NSF’s LIGO Laboratory which is a major facility fully funded by the National Science Foundation. This work is supported through Australian Research Council (ARC) Centres of Excellence CE170100004, CE230100016, Discovery Projects DP220101610 and DP230103088, and LIEF Project LE210100002. This work was performed on the OzSTAR national facility at Swinburne University of

Technology. The OzSTAR program receives funding in part from the Astronomy National Collaborative Research Infrastructure Strategy (NCRIS) allocation provided by the Australian Government, and from the Victorian Higher Education State Investment Fund (VHESIF) provided by the Victorian Government.

DATA AVAILABILITY

The data that support the findings of this article are openly available [56].

APPENDIX A: BUILDING REDUCED ORDER MODELS WITH ROMBUS

ROMBUS is a publicly available Python package for building reduced order models for arbitrary functions. In this section we describe how to make ROMs of gravitational waveforms using ROMBUS. However, the steps outlined offer a generalized guide for building ROMs of any model. We also describe changes made to ROMBUS to allow for the efficiency construction of ROMs. Finally, we include descriptions on the use and implementation of a number of gravitational wave specific algorithms utilized throughout this work, which are contained in a UTILS.PY file. This code is publicly available on GitHub at [56,57].

The first step in building a ROM is to define the base model to be compressed, which must be saved in a .PY file. This model must be a subclass of the generic RombusModel class. Each model must define a domain, codomain,¹² and prior ranges for all model parameters. In this work, the domain is frequency, the codomain is the strain h , and the prior ranges are built over the 15 parameters that parametrize the IMRPhenomPv2_NRTidalv2 waveform approximant. We improve the flexibility of ROMBUS by allowing arbitrary keyword arguments to be passed to the user-defined model class when loaded with ROMBUS. This is necessary to provide, for example, the priors and synthetic interferometers used for evaluating the frequency dependent antenna factors in the detector response. To read in the user-defined model, we pass the model file and keyword arguments *KWARGS to

```
model = RombusModel.load(model.py,
    *kwargs)
```

The model can now be passed into the model-building elements of ROMBUS. First, though, we must build the downsampled domain and the training set.

We define the `utils.down_sample` method to construct the downsampled domain based on the algorithm described in Sec. II D. The method takes the .INI file used for BILBY inference as an argument and reads the duration, frequency range, minimum component masses, and spins to

resample the frequency domain. It also takes an optional argument `df_max`, which defines the maximum frequency resolution for down sampling. This is important to prevent errors accruing from the time-to-merger approximation breaking down at high frequencies. We downsample the model domain by setting

```
model.domain = utils.down_sample
(filename.ini, df_max)
```

After setting the frequency domain, the training set is constructed. We instantiate the training set samples using

```
samples = Samples(model = model,
    n_random = 1)
```

Here, `SAMPLES` is the base ROMBUS class to hold parameter samples. We populate the samples with a single sample as specified by `n_random=1`. We define the methods `utils.boundary_samples`, `utils.grid_samples`, and `utils.random_samples` to populate the training set based on the methods outlined in Sec. II D. Only `model` and `n_samples`, an integer for the number of samples, need to be passed to the sampling methods.

The first method provides all combinations of the minimum and maximum prior boundaries to include in the training set. The second and third method populate the training set with `n_samples` samples taken from a uniform grid or randomly drawn from parameter space. The only parameter sampled from a nonuniform grid is the chirp mass $\mathcal{M}_c$, which is sampled uniformly in $\mathcal{M}_c^{-5/3}$ as this term describes the inspiral to leading order. This nonuniform sampling is also implemented in the `utils.grid_samples` method.

After obtaining the grid, random, and boundary samples, they are passed to the `samples` object by executing, for example,

```
samples.extend(boundary_samples)
```

To enable efficient parallelization, we rewrote the `extend` method to scatter the samples across all CPUs used to build the training set and keep track of how many samples are on each CPU. To obtain an equal number of samples on each CPU requires adding at most one extra sample to each CPU, which is computationally negligible.

Finally, the reduced order model is constructed by evaluating

```
rom = ReducedOrderModel(model, sam-
    ples, psd = psd).build(do_step=None, tol
    = tol)
```

Here, `model` is the `RombusModel` instance that describes our full order model, `samples` is a `samples` instance that we previously populated with parameter samples and is used to build the training set, and `PSD` is the set of weights passed to the noise-weighted inner product. In the case of gravitational-wave astronomy, we set the weights to the power spectral density of the detector network. The `ReducedOrderModel` class is instantiated with these arguments and the `build` method is called.

¹²In mathematics, a codomain is a set that includes all the possible values of a function—in this case, a set of predicted strain values.

Setting `do_step=None` tells Rombus to construct the reduced basis and the empirical interpolant. However, to construct one or the other simply set `do_step="RB"` or `do_step="EI"`, respectively.

The keyword argument `tol` is the user-defined tolerance for the maximum reduced basis approximation error. Throughout this work, we set the tolerance to `tol=1e-16`. This tolerance is comparable to the numerical precision of a double precision floating point number, used to represent the real and imaginary parts of the complex numbers used in this work. Depending on the problem at hand, such extreme levels of precision are not necessary.

In order to build almost any accurate reduced order model for gravitational-wave science, the reduced order modeling algorithm must be parallelized. A major contribution we have made to ROMBUS is parallelizing the reduced order modeling process, allowing for the efficient construction of ROMs on high performance computing clusters. Typically, we parallelize the ROM construction over hundreds of CPUs.

Calling `build(do_step=None, tol=tol)` create the training set out of the samples provided in `samples`, use the training set to construct the reduced basis, and then use the reduced basis to construct the empirical interpolant. These three data structures are contained in the ROM variable. So are the greedily selected parameters `rom.reduced_basis.greedy_points`. These are the parameters used to obtain the waveforms in the reduced basis.

As described in Sec. II D, these greedily selected parameters are integral to the up-sampling procedure of the reduced order model. Both the up-sampled linear and quadratic reduced order models are constructed over a training set consisting of *only* the greedily selected parameter samples. Even if adaptive sampling is not used, the quadratic model is still only constructed over these samples. This technique allows the user to dodge the extreme computational expense that comes with making a large training set over a uniform frequency domain.

In some cases, it is computationally impossible to make training sets large enough despite the adaptive sampling algorithm. In this case, we can increase the effective training set size—and therefore ROM accuracy—using the functionality

```
rom = rom.refine(N, tol=tol, iterate=False)
```

The `rom.refine` method tells ROMBUS to create a new training set of `N` random parameter samples and to rerun the reduced basis algorithm on the existing basis but using this new training set. This has the effect of increasing the effective training set size by `N` samples, as described in Sec. II D.

The keyword argument `iterate` dictates whether this refinement procedure proceeds until a completely new training set of `N` waveforms has no elements with approximation errors greater than `TOL`. Setting `iterate=True`

can cause the technique to iterate forever if using a tolerance comparable to machine precision due to numerical errors in the projection coefficients. Therefore, as we frequently use tolerances at this scale, we set `iterate=False` and loop over the refinement procedure `n_refine=3` times.

The components of the reduced order model that are actually used during inference are the interpolation matrix $B(f)$ and frequency nodes $\{F_i^{L/Q}\}_{i=1}^{N_{L/Q}}$ described in Sec. II A. These quantities can be saved to an output directory for use in inference. They can be easily accessed as

```
rom.empirical_interpolant.B_matrix
rom.empirical_interpolant.nodes
```

To use a reduced order model for inference using BILBY, simply add the following lines to an .INI file:

```
likelihood-type = ROQGravitationalWaveTransient
roq_folder = ROQ_data
```

Here, `ROQGravitationalWaveTransient` is the standard reduced order quadrature likelihood implemented in BILBY.

The reduced order quadrature likelihood used in this work [Eq. (17)] has been implemented in a forked version of BILBY, and is accessed by instead specifying

```
likelihood-type = TimeDependentROQGravitationalWaveTransient
```

The key difference in our implementation is that it allows the inclusion of the effects due to Earth’s rotation and the FSR, which is important to model for next-generation detectors. The standard `ROQGravitationalWaveTransient` likelihood implemented in BILBY does not include the effects due to Earth’s rotation and the FSR.

APPENDIX B: EFFECTS DUE TO EARTH’S ROTATION AND THE FREE SPECTRAL RANGE

Current ground-based gravitational wave detectors have arm lengths (L), which are much shorter than the wavelengths of the gravitational waves they are sensitive to (θ_{GW}). Equivalently, the characteristic frequency for light’s round trip along the arm, known as the free spectral range, $f_{fsr} = c/2L \approx 37 \text{ kHz} \gg f_{GW}$. In this limit (the “long wavelength approximation”), f_{GW} is approximately stationary along the light travel timescale, and the response of the detector to the signal is independent of f_{GW} .

With $L = 40 \text{ km}$, Cosmic Explorer (CE) will have an $f_{fsr} \approx 3.7 \text{ kHz}$, which is much closer to the gravitational wave frequencies it will be sensitive to. Relaxing the long-wavelength approximation, we must now account for the fact that the change in light travel time along a detector arm has additional dependence on how many cycles of the gravitational wave the light interacts with (previously it was assumed that the light was affected by the same phase of the gravitational wave signal, if present, for the entire duration of its trip down the arms). Accounting for this therefore

requires making the detector's response function explicitly depend on f_{GW} and the projection of the direction to the source onto the arm (see, e.g., [19,20]),

$$D(f, n_a) = \frac{f_{\text{fsr}}}{4\pi i f} \left(\frac{1 - e^{-\frac{\pi i f (1 - n_a)}{f_{\text{fsr}}}}}{1 - n_a} - e^{-\frac{2\pi i f}{f_{\text{fsr}}}} \frac{1 - e^{\frac{\pi i (1 + n_a) f}{f_{\text{fsr}}}}}{1 + n_a} \right), \quad (\text{B1})$$

where $n_a \equiv \vec{n} \cdot \hat{e}_a$ is the line-of-sight vector to the source $\vec{n}$ projected onto the unit vector $\hat{e}_a$ which points along arm a .

This form of D is used in Eq. (23) to compute $F^{+, \times}$. In the long-wavelength approximation used in typical analyses, D is a constant independent of f and $\vec{n}$.

The duration of signals seen in current ground-based detectors is much shorter than the timescale on which $\vec{n}$ changes due to the rotation of Earth, making it safe to fix $\vec{n}$ when projecting a simulated signal into the detector. With 3G detectors seeing signals for longer durations, these two timescales become comparable. We must therefore account for how the variation of $\vec{n}$ while the signal is visible affects the sensitivity of the detector to the signal.

While the equatorial/celestial coordinates (i.e., right ascension and declination) of $\vec{n}$ are fixed, Eq. (23) must be evaluated in the frame of the detector, a coordinate system in which $\vec{n}$ now explicitly depends on the current angular position between the detector and the source, making Eq. (23) depend on the current time. For long duration signals, this change in $\vec{n}$ means D and ϵ also change while the signal is in band. Because Eq. (23) is being evaluated on a grid of frequencies and $\vec{n}$ depends on the GPS time at the detector, we must calculate the time in the detector for each frequency point on the grid. In practice, we use a 1.5 PN approximation for the time to merger to calculate $t(f) - t_c$ for each point in the frequency grid f (e.g., [23]).

In BILBY, the signal at the detector location is calculated by first calculating $h(f)$ in the geocenter frame and then applying a time/phase shift to project the signal from the geocenter to the detector. This projection depends on the angle between $\vec{n}$ and the vector connecting the geocenter to the detector location. As this is a time-dependent projection for long duration signals, we correct for this Doppler effect in the phase by performing a time-dependent phase shift in the frequency domain,

$$h(f) = h_{\text{geocent}}(f) e^{-2\pi i f \delta t(f)}, \quad (\text{B2})$$

where $\delta t(f)$ is the difference in time between when the part of the signal with frequency f arrives in the geocenter and when it arrives at the detector. Whereas for analyses with current ground-based detectors, δt is a constant independent of f , δt is now an array with components determined by the time-to-merger approximation mentioned above.

APPENDIX C: PLOTS AND TABLES

This appendix includes Figs. 2–6, which provide supporting details to the main body. Figures 2–4 show the posterior distributions from inference performed on a GW170817-like binary neutron star signal injected into a single Cosmic Explorer with $\text{SNR} = 792$ and from a minimum frequency of $f_{\text{min}} = 26$ Hz. We use an ROQ rule that includes the effects due to Earth's rotation and the free spectral range. The signal was injected into a zero-noise realization of Gaussian noise, and therefore we expect the posterior distributions to peak at the injected parameters. In Fig. 5 we plot the real parts of the frequency-dependent and frequency-independent antenna response factors and phase of the detector response. In Fig. 6 we plot the maximum reduced basis approximation error as a function of the number of basis elements over a training set of 2 million binary neutron star waveforms starting from $f_{\text{min}} = 15$ Hz. Finally, in Table I we describe the size and accuracy of the reduced order models plotted in Fig. 1.

APPENDIX D: ANALYSIS TECHNIQUES

This appendix describes two useful results that support the main body. First, we derive an approximate relation between the maximum SNR of a signal that can be studied without introducing biases and the relative error in the log-likelihood ratio between the reduced order quadrature and standard likelihoods. Then, we demonstrate that it is computationally inefficient to construct ROMs for individual components of the detector response as opposed to a single ROM for the entire detector response.

1. Approximate *a priori* error bounds

We derive a relation between the relative log-likelihood ratio error and the maximum signal to noise ratio at which we can obtain unbiased inference. From [58], for unbiased parameter estimation we require $|h - h^{\text{interp}}|^2 \leq 1$. Furthermore, the log-likelihood ratio is on the order of the squared signal-to-noise ratio $|h|^2 = \langle h, h \rangle = \rho^2$ [21]. Therefore, the absolute relative error in log-likelihood ratio to leading order is given by

$$\left| \frac{\delta \ln \mathcal{L}}{\ln \mathcal{L}} \right| \approx \frac{||h|^2 - |h^{\text{interp}}|^2|}{|h|^2}. \quad (\text{D1})$$

Here, $\delta \ln \mathcal{L} = \ln \mathcal{L} - \ln \mathcal{L}_{\text{ROQ}}$. We can rewrite Eq. (D1) as

$$\left| \frac{\delta \ln \mathcal{L}}{\ln \mathcal{L}} \right| = \frac{||h|^2 + |h^{\text{interp}}|^2 - 2|h^{\text{interp}}|^2|}{|h|^2}. \quad (\text{D2})$$

If we assume that $|h| \approx |h^{\text{interp}}|$ (but $|h|^2 \not\approx |h^{\text{interp}}|^2$), and substitute $|h|^2 = \rho^2$ in the denominator, then

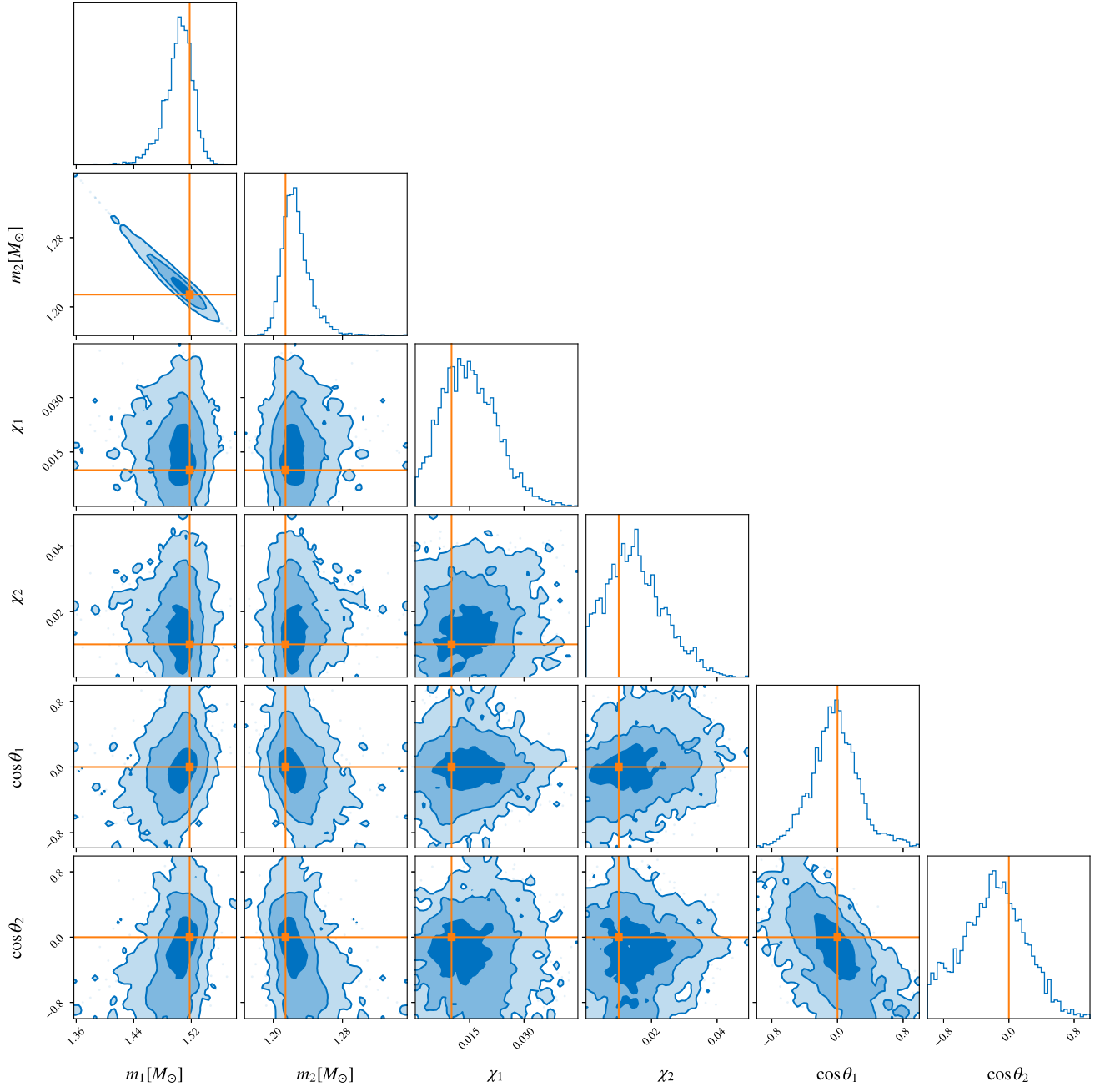


FIG. 2. One- and two-dimensional posterior distributions for the component masses $m_{1,2}[M_\odot]$, spin magnitudes $\chi_{1,2}$, and cosine of the spin tilts $\cos \theta_{1,2}$ of a GW170817-like binary neutron star signal injected into a single Cosmic Explorer with $\text{SNR} = 792$ and $f_{\min} = 26$ Hz. The true values are indicated in orange.

$$\left| \frac{\delta \ln \mathcal{L}}{\ln \mathcal{L}} \right| \approx \frac{||h|^2 + |h^{\text{interp}}|^2 - 2|h^{\text{interp}}||h||}{\rho^2}. \quad (\text{D3})$$

$$\leq \frac{1}{\rho^2}. \quad (\text{D5})$$

From the triangle inequality $||a| - |b|| \leq |a - b|$, we have that $|a|^2 + |b|^2 - 2|a||b| \leq |a - b|^2$. Therefore,

$$\left| \frac{\delta \ln \mathcal{L}}{\ln \mathcal{L}} \right| \leq \frac{|h - h^{\text{interp}}|^2}{\rho^2}, \quad (\text{D4})$$

In the last step, we have made use of the requirement that $|h - h^{\text{interp}}|^2 \leq 1$ from [58]. We wish to clarify that this requirement does hold for the reduced order models used in this work.

We claim that $|h - h^{\text{interp}}|^2 \leq 1$, and for now assume that $\langle h^{\text{interp}}, h^{\text{interp}} \rangle \approx \langle h, h \rangle = \rho^2$. Then,

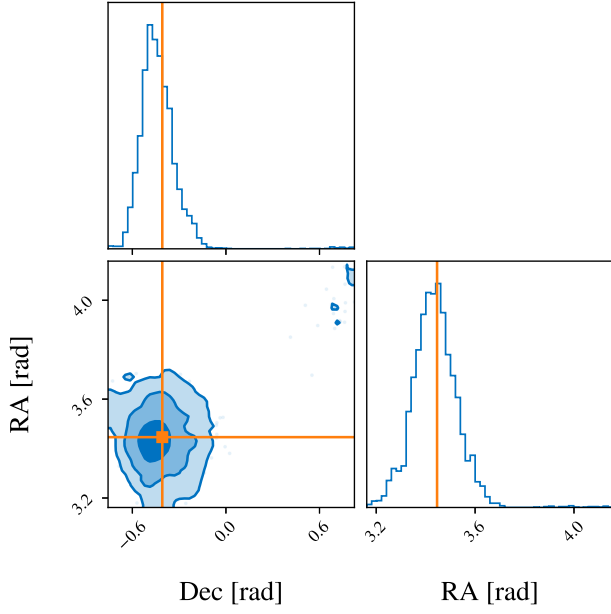


FIG. 3. One- and two- dimensional posterior distributions for the right ascension RA[rad] and declination Dec[rad] of a GW170817-like binary neutron star signal injected into a single Cosmic Explorer with SNR = 792 and $f_{\min} = 26$ Hz. The true values are indicated in orange.

$$|h - h^{\text{interp}}|^2 = \langle h - h^{\text{interp}}, h - h^{\text{interp}} \rangle, \quad (\text{D6})$$

$$= \langle h, h \rangle + \langle h^{\text{interp}}, h^{\text{interp}} \rangle - 2\langle h, h^{\text{interp}} \rangle, \quad (\text{D7})$$

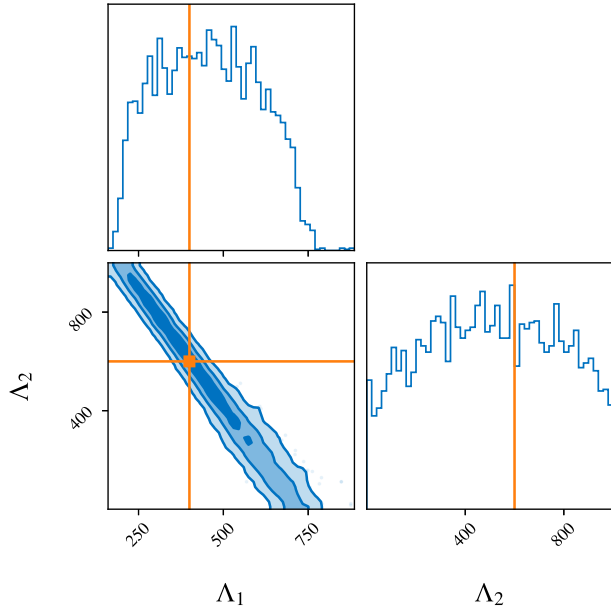


FIG. 4. One- and two- dimensional posterior distributions for the tidal deformabilities $\Lambda_{1,2}$ of a GW170817-like binary neutron star signal injected into a single Cosmic Explorer with SNR = 792 and $f_{\min} = 26$ Hz. The true values are indicated in orange.

$$\approx 2(\rho^2 - \langle h, h^{\text{interp}} \rangle). \quad (\text{D8})$$

The mismatch used throughout this work is defined as

$$M = 1 - \frac{\langle h, h^{\text{interp}} \rangle}{\sqrt{\langle h, h \rangle \langle h^{\text{interp}}, h^{\text{interp}} \rangle}}, \quad (\text{D9})$$

$$\approx 1 - \frac{\langle h, h^{\text{interp}} \rangle}{\rho^2}. \quad (\text{D10})$$

Therefore

$$\langle h, h^{\text{interp}} \rangle = \rho^2(1 - M). \quad (\text{D11})$$

Combining Eqs. (D8) and (D11) we arrive at

$$|h - h^{\text{interp}}|^2 \approx 2(\rho^2 - \rho^2 + \rho^2 M) = 2\rho^2 M. \quad (\text{D12})$$

The peak of the waveform mismatch for the reduced order models used throughout this work is typically $M \approx 10^{-12}$. If the typical signal-to-noise ratio is $\rho \approx 2 \times 10^3$, then

$$|h - h^{\text{interp}}|^2 \approx 10^{-5} \ll 1. \quad (\text{D13})$$

Therefore, from Eq. (D5) we see that the optimal signal-to-noise ratio sets an approximate upper bound on the relative error in the log-likelihood ratio required for unbiased inference.

2. Choice of reduced order quadrature

The greedy nature of the reduced basis algorithm implies that building separate bases for the components of the detector response will not yield greater compression than simply building a basis for the detector response itself. This is easily illustrated by considering separate models for the components of the phase-independent response, namely $h^{+/\times}$ and $F^{+/\times}$. Assuming these reduced order models approximate $F^\times$ and $h^\times$ well, then we have

$$F^{+/\times}(f) = \sum_a F^{+/\times}(F_a) A_a, \quad (\text{D14})$$

$$h^{+/\times}(f) = \sum_b h^{+/\times}(F_b) B_b. \quad (\text{D15})$$

The log-likelihood ratio is then

$$\ln \mathcal{L} = -\langle F^+ h^+, F^\times h^\times \rangle + \sum_{p=+/\times} \left(\langle d, F^p h^p \rangle - \frac{1}{2} \langle F^p h^p, F^p h^p \rangle \right). \quad (\text{D16})$$

If we consider the simplest term, $\langle d, F^+ h^+ \rangle$, we see that

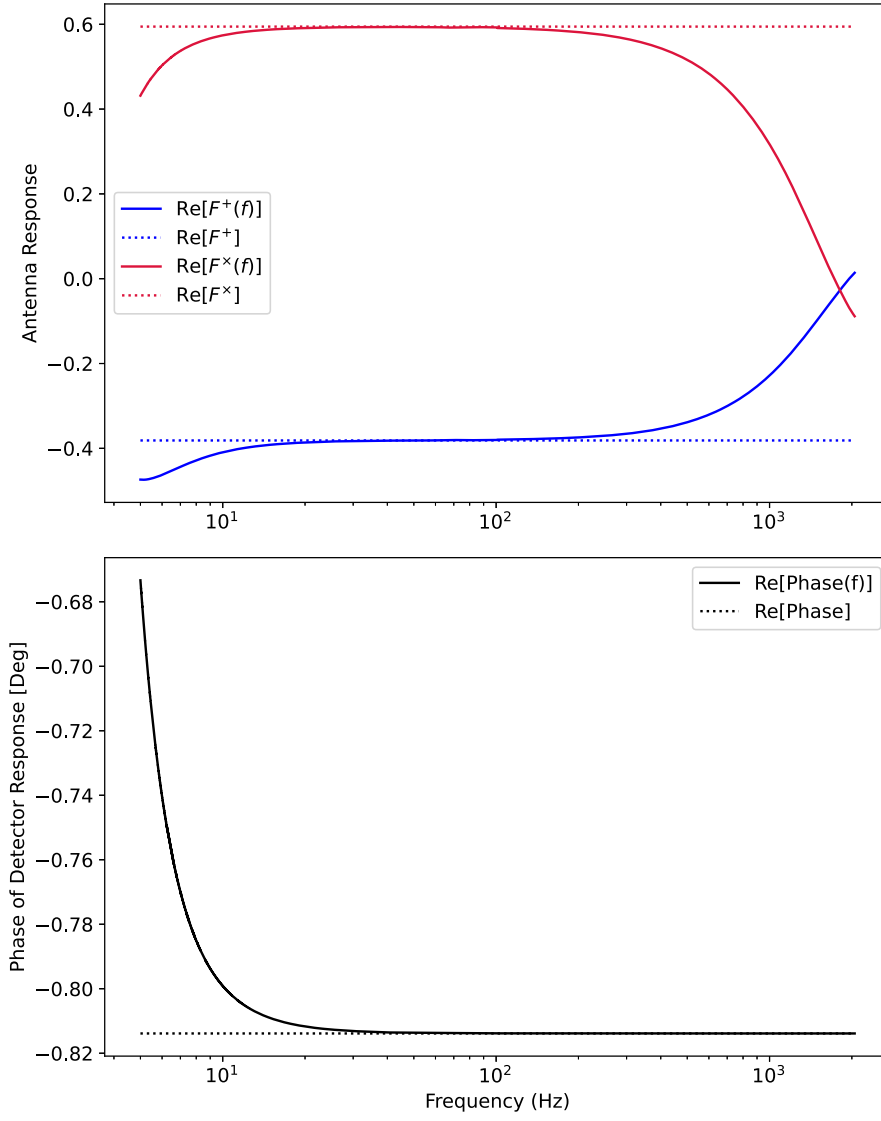


FIG. 5. Top: The real components of the plus- and cross-polarized antenna response factors in Eq. (23). Bottom: The real component of the phase of the detector response as defined in Eq. (8). All quantities are evaluated using Cosmic Explorer, a fixed source location at $(\text{dec}, \text{RA}) = (0, 0)$ and a polarization angle $\psi = 0$, and illustrated as frequency dependent (solid line) and frequency independent (dashed line).

$$\langle d, F^+ h^+ \rangle = \sum_a \sum_b \sum_i \frac{d_i F^+(F_a) A_a(f_i) h^+(F_b) B_b(f_i)}{P_i} \quad (\text{D17})$$

$$= \sum_a \sum_b F^+(F_a) h^+(F_b) \psi_{ab}. \quad (\text{D18})$$

Here, d_i is the gravitational wave data, P_i is the power spectral density of the detector, and $\{f_i\}_i^N$ is the discrete frequency domain over which the inference takes place. Furthermore, ψ_{ab} are the weights

$$\psi_{ab} = \sum_i \frac{d_i A_a(f_i) B_b(f_i)}{P_i}. \quad (\text{D19})$$

It is obvious that the same nested sum emerges for the remaining four sums in the log-likelihood ratio. Therefore, the log-likelihood ratio would scale as $5ab$. Given the greedy nature of the reduced basis algorithm, we can assume $ab \geq N_L$, where N_L is the size of the linear reduced basis required to represent $F^p h^p|_{p=+/\times}$. We can justify this as follows.

Consider the basis elements required to represent F^+ , given by $A = \{A_i\}_{i=1}^a$, and the basis elements required to represent h^+ , given by $B = \{B_j\}_{j=1}^b$. If F^+ and h^+ share no

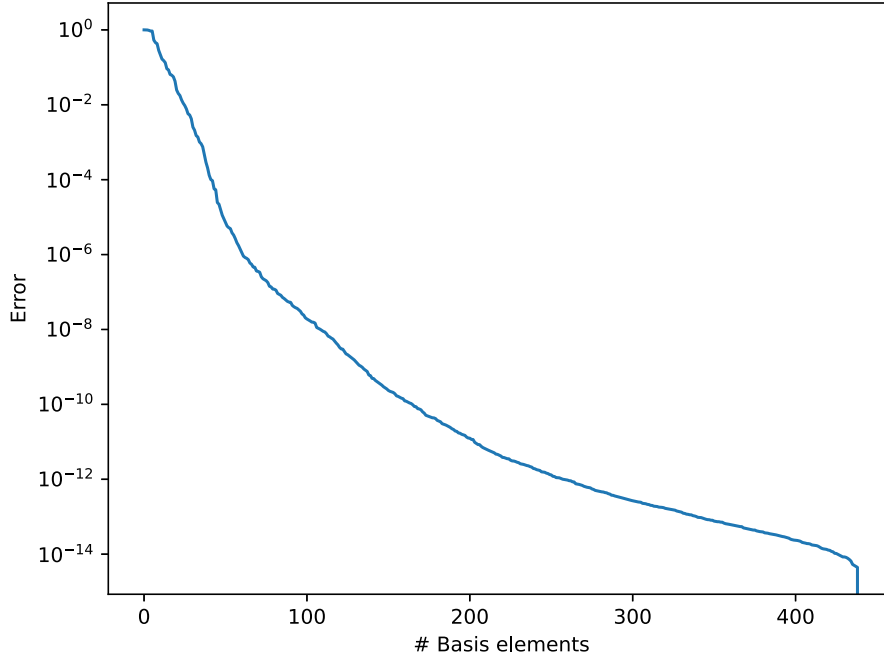


FIG. 6. Maximum reduced basis approximation error onto a training set of 2 million binary neutron star merger signals as a function of the number of basis elements. The signals are constructed with a minimum frequency of $f_{\min} = 15$ Hz. The tolerance for the maximum approximation error is $\epsilon = 10^{-16}$. At this error scale, numerical precision errors dominate and the mismatch between waveforms can become negative. This is why we observe a sharp drop at the 460th basis element: it is computationally infeasible to make a basis that is any more accurate.

TABLE I. Summary of ROMs in Fig. 1. Note that the ROM from $f_{\min} = 15$ Hz is built using a relaxed error tolerance in order to keep the basis size under 100 Gb, whereas the other ROMs are built using the same error tolerance of $\epsilon = 10^{-16}$.

$f_{\min}$ (Hz)	# Linear basis elements	Max SNR	$\Delta\mathcal{M}_c$ ($M_{\odot}$)
15	2029	608	2.2×10^{-3}
22	3388	1723	1×10^{-3}
24	3260	1793	9×10^{-4}
26	2666	1928	7.5×10^{-4}
28	2520	1973	7×10^{-4}
30	2493	2157	6×10^{-4}

information, then F^+ is not in the span of B . Then the basis required to represent F^+h^+ , given by $C = \{C_k\}_{k=1}^c$ with cardinality N_L , is simply

$$C = \{\alpha\beta | \alpha \in A, \beta \in B\}, \quad (\text{D20})$$

and therefore $N_L = ab$. However, if F^+ and h^+ share information, i.e., F^+ is in the span of B for some parameters θ (as $F^+ \equiv F^+(\theta)$), then

$$C \subseteq \{\alpha\beta | \alpha \in A, \beta \in B\}, \quad (\text{D21})$$

and has cardinality $N_L \leq ab$. This is a result of the spans of A and B overlapping.

The size of the quadratic basis is in general smaller than the linear basis as we no longer model the detector response phase. Therefore, we have $N_Q \leq N_L$ and we arrive at $5ab > 2N_L \geq N_L + N_Q$. So, constructing bases for components of the detector response is computationally inefficient compared to a single basis for the entire detector response.

- [1] M. Punturo, M. Abernathy, F. Acernese, B. Allen, and N. Andersson, The Einstein telescope: A third-generation gravitational wave observatory, *Classical Quantum Gravity* **27**, 194002 (2010).
- [2] D. Reitze *et al.*, Cosmic explorer: The U.S. contribution to gravitational-wave astronomy beyond LIGO, [arXiv:1907.04833](#).
- [3] E. D. Hall, K. Kuns, J. R. Smith, Y. Bai, and C. Wipf, Gravitational-wave physics with cosmic explorer: Limits to low-frequency sensitivity, *Phys. Rev. D* **103**, 122004 (2021).
- [4] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, and C. Affeldt (LIGO Scientific and Virgo Collaborations), GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs, *Phys. Rev. X* **9**, 031040 (2019).
- [5] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, and V. B. Adya (LIGO Scientific Collaboration and Virgo Collaboration), GWTC-2: Compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run, *Phys. Rev. X* **11**, 021053 (2021).
- [6] R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, N. Adhikari, R. X. Adhikari, and V. B. Adya (LIGO Scientific, Virgo, and KAGRA Collaborations), GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, *Phys. Rev. X* **13**, 041039 (2023).
- [7] J. E. Greene, J. Strader, and L. C. Ho, Intermediate-mass black holes, *Annu. Rev. Astron. Astrophys.* **58**, 259 (2020).
- [8] P. Baral, S. Morisaki, I. M. n. Hernandez, and J. Creighton, Localization of binary neutron star mergers with a single cosmic explorer, *Phys. Rev. D* **108**, 043010 (2023).
- [9] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, and F. Acernese, GW190425: Observation of a compact binary coalescence with total mass 3.4 m, *Astrophys. J. Lett.* **892**, L3 (2020).
- [10] J. W. Gardner, L. Sun, S. Borhanian, P. D. Lasky, E. Thrane, D. E. McClelland, and B. J. J. Slagmolen, Multi-messenger astronomy with a southern-hemisphere gravitational-wave observatory, *Phys. Rev. D* **108**, 123026 (2023).
- [11] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, and K. Ackley, Gw170817: Observation of gravitational waves from a binary neutron star inspiral, *Phys. Rev. Lett.* **119**, 161101 (2017).
- [12] E. Thrane and C. Talbot, An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models, *Pub. Astron. Soc. Aust.* **36**, e010 (2019).
- [13] Q. Hu and J. Veitch, Costs of bayesian parameter estimation in third-generation gravitational wave detectors: A review of acceleration methods, *Phys. Rev. D* **112**, 084039 (2025).
- [14] P. Canizares, S. E. Field, J. R. Gair, and M. Tiglio, Gravitational wave parameter estimation with compressed likelihood evaluations, *Phys. Rev. D* **87**, 124005 (2013).
- [15] S. Morisaki, Accelerating parameter estimation of gravitational waves from compact binary coalescence using adaptive frequency resolutions, *Phys. Rev. D* **104**, 044062 (2021).
- [16] N. J. Cornish, Fast fisher matrices and lazy likelihoods, [arXiv:1007.4820](#).
- [17] N. Leslie, L. Dai, and G. Pratten, Mode-by-mode relative binning: Fast likelihood estimation for gravitational waveforms with spin-orbit precession and multiple harmonics, *Phys. Rev. D* **104**, 123030 (2021).
- [18] R. Smith, S. E. Field, K. Blackburn, C.-J. Haster, M. Pürrer, V. Raymond, and P. Schmidt, Fast and accurate inference on gravitational waves from precessing compact binaries, *Phys. Rev. D* **94**, 044031 (2016).
- [19] M. Rakhmanov, J. D. Romano, and J. T. Whelan, High-frequency corrections to the detector response and their effect on searches for gravitational waves, *Classical Quantum Gravity* **25**, 184017 (2008).
- [20] R. Essick, S. Vitale, and M. Evans, Frequency-dependent responses in third generation gravitational-wave detectors, *Phys. Rev. D* **96**, 084004 (2017).
- [21] S. Morisaki, R. Smith, L. Tsukada, S. Sachdev, S. Stevenson, C. Talbot, and A. Zimmerman, Rapid localization and inference on compact binary coalescences with the Advanced LIGO-Virgo-KAGRA gravitational-wave detector network, *Phys. Rev. D* **108**, 123040 (2023).
- [22] R. Smith, S. Borhanian, B. Sathyaprakash, F. Hernandez Vivanco, and S. E. Field, Bayesian inference for gravitational waves from binary neutron star mergers in third generation observatories, *Phys. Rev. Lett.* **127**, 081102 (2021).
- [23] C. Cutler and E. E. Flanagan, Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral waveform?, *Phys. Rev. D* **49**, 2658 (1994).
- [24] L. Santamaría *et al.*, Matching post-Newtonian and numerical relativity waveforms: Systematic errors and a new phenomenological model for nonprecessing black hole binaries, *Phys. Rev. D* **82**, 064016 (2010).
- [25] F. Iacovelli, M. Mancarella, S. Foffa, and M. Maggiore, Forecasting the detection capabilities of third-generation gravitational-wave detectors using GWFIRST, *Astrophys. J.* **941**, 208 (2022).
- [26] R. DeVore, G. Petrova, and P. Wojtaszczyk, Greedy algorithms for reduced bases in banach spaces, *Constr. Approx.* **37**, 455 (2013).
- [27] G. Rozza, D. Huynh, and A. Patera, Reduced basis approximation and *a posteriori* error estimation for affinely parametrized elliptic coercive partial differential equations, *Arch. Comput. Methods Eng.* **15**, 229 (2008).
- [28] M. Barrault, Y. Maday, N. C. Nguyen, and A. T. Patera, An empirical interpolation method: Application to efficient reduced-basis discretization of partial differential equations, *Rep. Math.* **339**, 667 (2004).
- [29] S. E. Field, C. R. Galley, J. S. Hesthaven, J. Kaye, and M. Tiglio, Fast prediction and evaluation of gravitational waveforms using surrogate models, *Phys. Rev. X* **4**, 031006 (2014).
- [30] K. Chatziioannou, T. Dent, M. Fishbach, F. Ohme, M. Pürrer, V. Raymond, and J. Veitch, Compact binary coalescences: Gravitational-wave astronomy with ground-based detectors, [arXiv:2409.02037](#).

- [31] B. Zackay, L. Dai, and T. Venumadhav, Relative binning and fast likelihood evaluation for gravitational wave parameter estimation, [arXiv:1806.08792](https://arxiv.org/abs/1806.08792).
- [32] T. Dietrich, A. Samajdar, S. Khan, N.K. Johnson-McDaniel, R. Dudi, and W. Tichy, Improving the NRTidal model for binary neutron star systems, *Phys. Rev. D* **100**, 044003 (2019).
- [33] J. S. Speagle, DYNESTY: A dynamic nested sampling package for estimating Bayesian posteriors and evidences, *Mon. Not. R. Astron. Soc.* **493**, 3132 (2020).
- [34] G. Ashton *et al.*, Bilby: A user-friendly Bayesian inference library for gravitational-wave astronomy, *Astrophys. J. Suppl. Ser.* **241**, 27 (2019).
- [35] I. M. Romero-Shaw *et al.*, Bayesian inference for compact binary coalescences with BILBY: Validation and application to the first LIGO–Virgo gravitational-wave transient catalogue, *Mon. Not. R. Astron. Soc.* **499**, 3295 (2020).
- [36] G. Ashton and C. Talbot, Bilby-MCMC: An MCMC sampler for gravitational-wave inference, *Mon. Not. R. Astron. Soc.* **507**, 2037 (2021).
- [37] H.-Y. Chen, P. S. Cowperthwaite, B. D. Metzger, and E. Berger, A program for multimessenger standard siren cosmology in the era of LIGO a+, rubin observatory, and beyond, *Astrophys. J. Lett.* **908**, L4 (2021).
- [38] M. Pürrer and C.-J. Haster, Gravitational waveform accuracy requirements for future ground-based detectors, *Phys. Rev. Res.* **2**, 023151 (2020).
- [39] I. M. Romero-Shaw, E. Thrane, and P. D. Lasky, When models fail: An introduction to posterior predictive checks and model misspecification in gravitational-wave astronomy, *Pub. Astron. Soc. Aust.* **39**, E025 (2022).
- [40] C. B. Owen, C.-J. Haster, S. Perkins, N. J. Cornish, and N. Yunes, Waveform accuracy and systematic uncertainties in current gravitational wave observations, *Phys. Rev. D* **108**, 044018 (2023).
- [41] B. P. Abbott *et al.*, A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals, *Classical Quantum Gravity* **37**, 055002 (2020).
- [42] C. Talbot and E. Thrane, Gravitational-wave astronomy with an uncertain noise power spectral density, *Phys. Rev. Res.* **2**, 043298 (2020).
- [43] S. Biscoveanu, C.-J. Haster, S. Vitale, and J. Davies, Quantifying the effect of power spectral density uncertainty on gravitational-wave parameter estimation for compact binary sources, *Phys. Rev. D* **102** (2020).
- [44] C. Talbot, E. Thrane, S. Biscoveanu, and R. Smith, Inference with finite time series: Observing the gravitational Universe through windows, *Phys. Rev. Res.* **3**, 043049 (2021).
- [45] A. G. e. a. Martini A. and Schmidt S., Maximum entropy spectral analysis: An application to gravitational waves data analysis., *Eur. Phys. J. C* **84**, 1023 (2024).
- [46] K. Chatziioannou, N. Cornish, M. Wijngaarden, and T. B. Littenberg, Modeling compact binary signals and instrumental glitches in gravitational wave data, *Phys. Rev. D* **103**, 044013 (2021).
- [47] C. Plunkett, S. Hourihane, and K. Chatziioannou, Concurrent estimation of noise and compact-binary signal parameters in gravitational-wave data, *Phys. Rev. D* **106**, 104021 (2022).
- [48] N. Guttman, P. D. Lasky, and E. Thrane, Modelling noise in gravitational-wave observatories with transdimensional models, *Phys. Rev. D* **111**, 063063 (2025).
- [49] R. Essick, Calibration uncertainty’s impact on gravitational-wave observations, *Phys. Rev. D* **105**, 082002 (2022).
- [50] B. P. Abbott *et al.*, GW170817: Implications for the stochastic gravitational-wave background from compact binary coalescences, *Phys. Rev. Lett.* **120**, 091101 (2018).
- [51] A. D. Johnson, K. Chatziioannou, and W. M. Farr, Source confusion from neutron star binaries in ground-based gravitational wave detectors is minimal, *Phys. Rev. D* **109**, 084015 (2024).
- [52] S. Biscoveanu, C. Talbot, E. Thrane, and R. Smith, Measuring the primordial gravitational-wave background in the presence of astrophysical foregrounds, *Phys. Rev. Lett.* **125**, 241101 (2020).
- [53] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, Real-time gravitational-wave science with neural posterior estimation, *Phys. Rev. Lett.* **127**, 241103 (2021).
- [54] M. Dax, S. R. Green, J. Gair, N. Gupte, M. Pürrer, V. Raymond, J. Wildberger, J. H. Macke, A. Buonanno, and B. Schölkopf, Real-time gravitational-wave inference for binary neutron stars using machine learning, *Nature (London)* **639**, 49 (2025).
- [55] Q. Hu, J. Irwin, Q. Sun, C. Messenger, L. Suleiman, I. S. Heng, and J. Veitch, Decoding long-duration gravitational waves from binary neutron stars with machine learning: Parameter estimation and equations of state, *Astrophys. J. Lett.* **987**, L17 (2025).
- [56] A. M. Baker, [10.5281/zenodo.17354925](https://doi.org/10.5281/zenodo.17354925) (2024), version 1.0.0.
- [57] <https://github.com/amakaib/gwrombus>
- [58] L. Lindblom, B. J. Owen, and D. A. Brown, Model waveform accuracy standards for gravitational wave data analysis, *Phys. Rev. D* **78**, 124020 (2008).