Combining effective-one-body accuracy and reduced-order-quadrature speed for binary neutron star merger parameter estimation with machine learning

Jacopo Tissino,[1,2] Gregorio Carullo,[3,4] Matteo Breschi,[3] Rossella Gamba,[3] Stefano Schmidt[5,6] and Sebastiano Bernuzzi[3]

[1]*Gran Sasso Science Institute (GSSI), I-67100 L'Aquila, Italy*
[2]*INFN, Laboratori Nazionali del Gran Sasso, I-67100 Assergi, Italy*
[3]*Theoretisch-Physikalisches Institut, Friedrich-Schiller-Universität Jena, 07743, Jena, Germany*
[4]*Niels Bohr International Academy, Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, Denmark*
[5]*Nikhef, Science Park 105, 1098 XG, Amsterdam, The Netherlands*
[6]*Institute for Gravitational and Subatomic Physics (GRASP), Utrecht University,
Princetonplein 1, 3584 CC Utrecht, The Netherlands*

We present `mlgw_bns`, a gravitational waveform surrogate that allows for a significant improvement in the generation speed of frequency-domain waveforms for binary neutron star mergers, at a negligible cost in accuracy. This improvement is achieved by training a machine-learning model on a dataset of waveforms generated with an accurate but comparatively costlier approximant: the state-of-the-art effective-one-body model `TEOBResumSPA`. When coupled to a reduced-order scheme, `mlgw_bns` can accelerate waveform generation up to a factor of ~ 35, outperforming all other approximants of similar accuracy. By analyzing GW170817 in realistic parameter estimation settings with our scheme, we showcase an overall speedup against `TEOBResumSPA` greater than an order of magnitude. Our methodology will bear a significant impact on the scientific program of next generation detectors by allowing routine usage of accurate effective-one-body models.

I. INTRODUCTION

Bayesian analyses of gravitational wave (GW) data from compact binary mergers rely on extensive explorations of the posterior probability distribution of detected signals [1,2] and chiefly hinge on accurate waveform models. The latter represent the prediction of a GW signal originated from a system described by a certain set of parameters θ. As the sampling of the posterior distribution for a single GW event typically requires the generation of $\gtrsim 10^7$ waveforms, speed in their generation is essential. This is especially compelling in view of next generation (XG) GW detectors, for which the rate of events will be dramatically higher than the current one. The in-band duration of signals will also increase and, due to their low mass, long signals emitted by binary neutron star (BNS) mergers will be most significantly affected. Reducing the computational cost for this class of signals, maximizing the scientific output of future large-scale experiments, is a goal of the utmost importance—this is the focus of the present work.

For general relativistic waveform models, speed and accuracy are often at odds. For example, very fast waveform generation can be obtained with analytical Post-Newtonian (PN) approximants [3,4], but such templates lack in accuracy and tend to bias parameter estimation (PE). Indeed, Bayesian analyses of BNS signals in XG detectors [5–7] have been demonstrated only using phenomenological [8] or PN approximants [9]. These waveform templates include only partial physical information, and are therefore expected to strongly bias PE with XG detectors [7,10]. For example, phenomenological approximants model the effect of spin precession, but do not contain unequal-mass tidal corrections, with the binary matter effects entirely determined by one single effective tidal parameter. Post-Newtonian models are unreliable close to merger and are only available (in the frequency domain) for binaries with spins aligned with the orbital angular momentum. Incorporating the whole physics content of advanced waveform models (including higher harmonics, precession, eccentricity, self-spin interactions, beyond leading order adiabatic electric and magnetic-type tidal effects, and dynamical tides) will be key to avoid biases, but currently can only be accomplished at a great increase in computational cost.

In particular, here we are interested in leveraging the effective-one-body (EOB) approach [11–20], one of the most accurate state-of-the art frameworks for waveform generators. In this framework, the Hamiltonian description of the two-body problem in General Relativity (GR) is

mapped to an effective problem of a single body orbiting in a Kerr-like deformed metric. The effective metric potentials are determined by suitably resummed PN expressions that make the model predictive in the fast-motion and strong-field merger regime. Gravitational waveforms are natively generated in the time domain using the solution of the EOB equations of motion and a particular factorized and resummed analytical expression of the multipolar PN waveform [21]. The EOB approach has the advantage of being both accurate to Einstein's equations, and flexible to the addition of analytical (e.g., PN) information. The faithfulness of inspiral-merger-ringdown models is increased by suitably informing them with Numerical Relativity (NR) data, see, e.g., Refs. [22,23] for recent work targeted at XG detectors. Analogously, BNS inspiral-merger waveforms are obtained by augmenting the effective interbinary potential and waveform multipoles with tidal terms [24–31]. Full inspiral-merger-postmerger BNS waveforms can be constructed by hybridizing the model with NR-informed post-merger models [32,33].

The practical usage of the above EOB model is hampered by the requirement of numerically solving an ODE system, which brings a constant-time overhead and constrains the maximum rate of waveform generation. For current BNS analyses, generating the $\gtrsim 10^7$ templates required easily takes several weeks of CPU time. A crucial element to improve the EOB model efficiency is the post-adiabatic method, introduced in Ref. [34] for the `TEOBResumS` model. The post-adiabatic iterative method yields an efficient yet accurate approximation of the EOB Hamiltonian flow, removing the need to solve the related ODE for all but the very last stages of the inspiral. This technique provides a significant speedup (a factor of ten or more for typical BNS signals in the LIGO-Virgo [35,36] band), but it is currently applicable only to quasicircular mergers. To further optimize the waveform generation, a desirable feature for a fast approximant is to yield waveforms in the frequency domain, since the likelihood takes a simple form in the Fourier space, when assuming a Gaussian and stationary noise background. A time-domain approximant, such as the one mentioned above, needs to be Fourier-transformed before use, which typically entails a slowing-down up to an order of magnitude. For this reason, a stationary phase approximation (SPA) was introduced within the EOB model `TEOBResumS`, yielding a fast and accurate frequency-domain approximant called `TEOBResumSPA` [37], which has been successfully applied to the analyses of GW170817 and GW190425 data, see, e.g., Refs [37,38].

The evaluation of a frequency-domain waveform approximant typically scales as $t_{\text{waveform}} \approx t_{\text{overhead}} + N_{\text{points}} t_{\text{point}}$, where N_{points} is the number of grid points it is evaluated at. The per-frequency-point time t_{point} is typically on the order of few hundreds of nanoseconds, cannot be reduced below the CPU clock speed times the number of floating point operations required, and varies much less than t_{overhead} across

models. Thanks to the combination of SPA and the post-adiabatic approach, the overhead time for `TEOBResumSPA` has been reduced to only tens of milliseconds. The fundamental limitation in reducing this number is that, even when evaluating the waveform at few frequency points, the full Hamiltonian flow must still be computed in a complete radial grid. In current BNS analyses the second term is typically dominant. This leads to PE times on the order of a few days on a modern computer cluster, which is acceptable for current event rates, but will not be for XG detectors. A key observation is that Reduced Order Quadratures (ROQs) [39] techniques can decrease the required value of N_{points} so much that the linear term becomes negligible compared to the constant one. The driving requirement behind this work is therefore to build a model with a much lower t_{overhead}—which leads to a significantly faster PE if combined with ROQ—while remaining faithful to the predictions of EOB; this will enable accurate analyses of data from XG detectors.

One of the most promising approaches to achieving the necessary increase in efficiency is template acceleration through machine learning (ML). The last few years saw a sharp rise in studies on this topic, a review of which can be found in Ref. [40]. Most of these efforts, however, focused on binary black hole (BBH) signals. A pioneering study on this was the one of Ref. [41], which developed a neural network to compute the liner combination coefficients of a generic BBH represented on a basis of waveforms. Along the lines of this work, Ref. [42] also reached high performance, while retaining faithfulness, compared the training waveforms. Reference [43] used instead a principal component analysis (PCA) to drastically reduce the number of basis functions required for waveform reconstruction, while Ref. [44] applied automated learning to select the best performing regression scheme (although varying only the BBH mass ratio q) and Ref. [45] extended the latter effort to spin-precessing signals. Finally, Ref. [46] used a deep generative model for waveform generation. It is worth noting that the models listed above work in the time-domain; while this ensures a smooth (hence more easily learnable) physical representation of the BBH signal, a Fourier transform is still required to use the model in PE applications. The literature is less rich in the field of binary neutron star (BNS) modeling. Reference [47] developed a nonspinning surrogate model in the time domain, and extended it to aligned-spin BNS using Gaussian process regression [20]. Their parameter space is the same one we use, as discussed later. Subsequently, Ref. [20] built a fast frequency-domain surrogate of the spin-aligned model `SEOBNRv4T` [29,30], again using Gaussian process regression.

In this work, we boost the efficiency of accurate EOB models by introducing `mlgw_bns`, the first *frequency domain* BNS surrogate model relying on a neural network. The key advancements introduced by our training algorithm rely on a combination of data compression

techniques and analytical knowledge exploitation. Its salient characteristics are: training on `TEOBResumSPA` waveform residuals relative to a PN baseline, multistep waveform downsampling, and a final PCA compression stage. This way, the neural network must only learn a much simplified relation between the BNS parameters θ and a low-dimensional representation of the waveform, which allows it to be shallow, in the end significantly decreasing the waveform computational overhead. Synergic usage of this model with ROQ compression techniques allows us to showcase more than an order of magnitude improvement in the analysis of current BNS signals. Most importantly, our technique immediately provides speedups up to $\sim 10^3$ for wider bandwidth analyses, enabling future studies to systematically exploit highly accurate EOB models in full Bayesian PE analyses involving XG detectors.

This paper is organized as follows. In Sec. II, we describe the details of our method, while Sec. III is devoted to the performance analysis of our model in terms of timing and accuracy. The improved capabilities of the resulting model are illustrated in Sec. IV, where we show the results of a realistic PE analysis on the BNS transient GW170817 [48,49], additionally making use of a reduced order quadrature scheme to fully exploit the potential of our technique. Section V presents final remarks and future research directions.

Software availability.—Our model is released within the public Python package `mlgw_bns`, available at pypi.org/project/mlgw-bns/. The description in this paper refers to version `0.12.0`. The package contains both the trained model described here, which can be used to generate waveforms out of the box, and the full functionalities required to train new models at will (e.g., by using a different approximant than `TEOBResumSPA`, or different parameter ranges). The training time and memory requirements are both relatively small: a model can easily be trained on a laptop in a few hours. The software we developed to achieve the frequency compression applied in the PE stage is available at: github.com/GCArullo/JenpyROQ.

Conventions.—We work in geometric units, setting $G = c = 1$. The total binary mass is denoted as $M = m_1 + m_2$, the mass ratio as $q = m_1/m_2 \geq 1$, and the symmetric mass ratio as $\nu = m_1 m_2/M^2$. The dimensionless spin vectors are denoted as χ_i for $i = 1, 2$ and the spin components aligned with the orbital angular momentum $\mathbf{L}$ are labeled as $\chi_i = \chi_i \cdot \mathbf{L}/|\mathbf{L}|$. The effective spin parameter is defined as $\chi_{\text{eff}} = (\chi_1 m_1 + \chi_2 m_2)/M$. The quadrupolar tidal polarizability parameters are defined as $\Lambda_i = (2/3)k_{2,i}C_i^{-5}$ for $i = 1, 2$, where $k_{2,i}$ and C_i are the second Love number and the compactness of the ith star, respectively. The reduced tidal parameter,

$$\tilde{\Lambda} = \frac{16}{13}\frac{(M_1 + 12M_2)M_1^4}{M^5}\Lambda_1 + (1 \leftrightarrow 2), \quad (1)$$

determines tidal interactions at leading post-Newtonian-order [50,51]. Masses, spins, and tidal parameters are collectively called the *intrinsic parameters* of a BNS system, i.e., $\theta_{\text{int}} = \{M, q, \chi_1, \chi_2, \Lambda_1, \Lambda_2\}$. The location and orientation of the source are identified by the *extrinsic parameters* $\theta_{\text{ext}} = \{D_L, \iota, \alpha, \delta, \psi, t_c, \phi_c\}$, i.e., luminosity distance D_L, inclination angle ι, right ascension angle α, declination angle δ, polarization angle ψ, time of coalescence t_c, and phase at the merger ϕ_c.[1]

The frequency-domain waveform from a compact binary coalescence can in general be written as

$$h_+(f) - ih_\times(f) = \frac{1}{D_L}\sum_{l=2}^{\infty}\sum_{m=-\ell}^{\ell} h_{\ell m}(f)_{(-2)}Y_{\ell m}(\iota, \varphi), \quad (2)$$

where the functions $_{(-2)}Y_{\ell m}$ are the spin-weighted spherical harmonics, given, e.g., by Equations (II.7) and (II.8) of Ref. [52], while the complex functions $h_{\ell m}(f)$ are the frequency-domain modes of the GW strain.

The discussion in this paper is restricted to the $(\ell, m) = (2, 2)$ mode; focusing on it, the two GW polarizations can be simply written as

$$h_+(f) = \frac{1}{D_L}\sqrt{\frac{5}{4\pi}}h_{22}(f)\frac{\cos^2\iota + 1}{2} \quad (3a)$$

$$h_\times(f) = \frac{1}{D_L}\sqrt{\frac{5}{4\pi}}h_{22}(f)\cos\iota. \quad (3b)$$

A relevant scalar product in waveform space is the Wiener product

$$(a|b) := 4\Re \int_0^\infty \frac{a^*(f)b(f)}{S_n(f)}\,df, \quad (4)$$

where S_n is the power spectral density (PSD) of a given detector. Results shown in Sec. III are computed considering S_n to be the expected Einstein Telescope PSD, ET-D [53,54].

In terms of this, the optimal match (or *faithfulness*) between waveforms a and b is given by

$$\mathcal{F}(a, b) := \max_{t_0, \phi_0} \frac{(a|b)}{\sqrt{(a|a)(b|b)}}, \quad (5)$$

where the maximum is taken over all possible time and phase shifts t_0 and ϕ_0 between the two waveforms. The mismatch is then defined as $\bar{\mathcal{F}}(a, b) := 1 - \mathcal{F}(a, b)$.

---

[1]This is the full extrinsic parameter set required to reconstruct the $h_{\mu\nu}$ tensor, but within `mlgw_bns` concretely the sky position parameters α, δ, and ψ are not accepted: the polarizations are returned in a frame located at Earth and aligned with the source.

II. MODEL CONSTRUCTION

A. Overview

`mlgw_bns` is a surrogate waveform approximant based on a neural network that learns the relation between five intrinsic parameters of a binary system—mass ratio, dimensionless spins, and quadrupolar tidal polarizabilities, collectively denoted as $\theta = \{q, \chi_1, \chi_2, \Lambda_1, \Lambda_2\}$—and the corresponding frequency-domain waveform mode $h_{22}(f; \theta)$. The binary mass M is not included in θ since the nontrivial mass scale in the BNS problem is fully included in the tidal polarizability parameters.[2] Concretely, within `mlgw_bns` a fixed reference mass of $M_{\text{ref}} = 2.8 M_\odot$ is chosen and waveforms for generic masses are generated by the appropriate rescaling of both the waveform's amplitude and frequency; this is described in more detail in Sec. II D. Similarly, the other extrinsic parameters θ_{ext} can be neglected when constructing an approximant; the dependence on them can be included analytically in the likelihood after a waveform has been generated.

The driving idea behind `mlgw_bns` is to have the neural network be as shallow and small as possible while retaining reconstruction accuracy; this is accomplished by reducing the dimensionality of the waveform's description. The first step to this end is to make a training dataset of *residuals* from an analytical PN baseline, which means the network only has to learn information in the high-frequency region, where the two models differ; this is described in Sec. II B.

We then employ a reduced frequency grid and perform a PCA in order to decrease the dimensionality of each waveform's representation to about 30 floating point numbers; this is described in Sec. II C.

Finally, a neural network is trained to reconstruct the relation between the parameters θ and the $\sim$30 principal components, as described in Sec. II E.

The training datasets for all the aforementioned stages are generated by drawing from the same uniform distribution on the parameters, in the intervals:

$$q \in [1, 2], \qquad \Lambda_i \in [5, 5000], \qquad \chi_i \in [-0.5, 0.5].$$

These ranges correspond to a realistic prior choice in GW analyses of BNS. The random number generator used for the extraction is deterministically reseeded for every new dataset, in order to ensure reproducibility as well as independence of the datasets.

The frequency-domain waveforms currently learned by `mlgw_bns` are those generated by the state-of-the-art EOB model `TEOBResumSPA`; these will be denoted by a subscript EOB in the following discussion. We train with `TEOBResumSPA` frequency-domain waveforms as opposed

---

[2]As in the scale-invariant binary black hole case, the waveform's frequency dependence is really on the mass-rescaled parameter $Mf = GMf/c^3$, not on f alone.

to Fourier transforms of time-domain `TEOBResumS` waveforms for a few reasons: the two are closer than the intrinsic accuracy of `TEOBResumS` ($\bar{\mathcal{F}} \lesssim 5 \times 10^{-4}$ [37]); the SPA waveforms are much smoother than the ones calculated with a fast Fourier transform (FFT), and therefore easier to represent with small amounts of frequency points (see Sec. II C 2); the SPA waveforms can be natively evaluated at arbitrary frequencies, allowing us to never employ a uniform frequency grid.

Figure 1 shows a graphical outline of `mlgw_bns`'s operation.

B. Residuals from a Post-Newtonian baseline

We start with a polar representation of the waveform in amplitude and phase as $h(f) = A_{\text{EOB}}(f)e^{-i\phi_{\text{EOB}}(f)}$. Instead of reconstructing the waveform directly, `mlgw_bns` reconstructs its residuals from a fiducial PN model. The residuals are computed as

$$\Delta A(f; \theta) = \log\left(\frac{A_{\text{EOB}}(f; \theta)}{A_{\text{PN}}(f; \theta)}\right) \tag{6a}$$

$$\Delta\phi(f; \theta) = \phi_{\text{EOB}}(f; \theta) - \phi_{\text{PN}}(f; \theta), \tag{6b}$$

and shown in Fig. 2 for 100 sets of parameters. The complete waveform is recovered from the predicted residuals $\Delta A_{\text{pred}}(f; \theta), \Delta\phi_{\text{pred}}(f; \theta)$ as

$$A_{\text{pred}}(f; \theta) = A_{\text{PN}}(f; \theta) \exp(\Delta A_{\text{pred}}(f; \theta)) \tag{7a}$$

$$\phi_{\text{pred}}(f; \theta) = \phi_{\text{PN}}(f; \theta) + \Delta\phi_{\text{pred}}(f; \theta). \tag{7b}$$

We use the `TaylorF2` approximant with 3.5PN-accurate amplitude, pseudo 5.5PN-accurate phase [55] with 7.5PN-accurate tidal contributions [51,56], and the monopole-quadrupole 3PN contribution to the phase [[57], Eqs. (50)–(52)] (see also [[19], Eq. (41)]).

The phase residuals computed as above typically exhibit large linear trends due to the different choices in the time-domain alignment between the EOB and PN models (which corresponds to a linear phase term in the frequency domain). These trends are not physically meaningful, but even small differences can result in a large effect; the variation over the whole frequency spectrum is of the order of 2000 Hz $\times$ $2\pi \times \Delta t$ radians (for the reference mass), meaning that even single millisecond shifts will yield tens of radians in difference. Typical shifts between the models used within `mlgw_bns` are of the order of tens of milliseconds, resulting in several hundreds of radians of meaningless phase difference. In order to remove this effect, the average slope $d\Delta\phi/df$ is first calculated between the first frequency sample and some higher frequency (typically chosen to be low enough to lie in the region of validity of the PN approximation) and then the corresponding linear term is subtracted from the residuals. Figure 2 shows residuals with this procedure already applied.
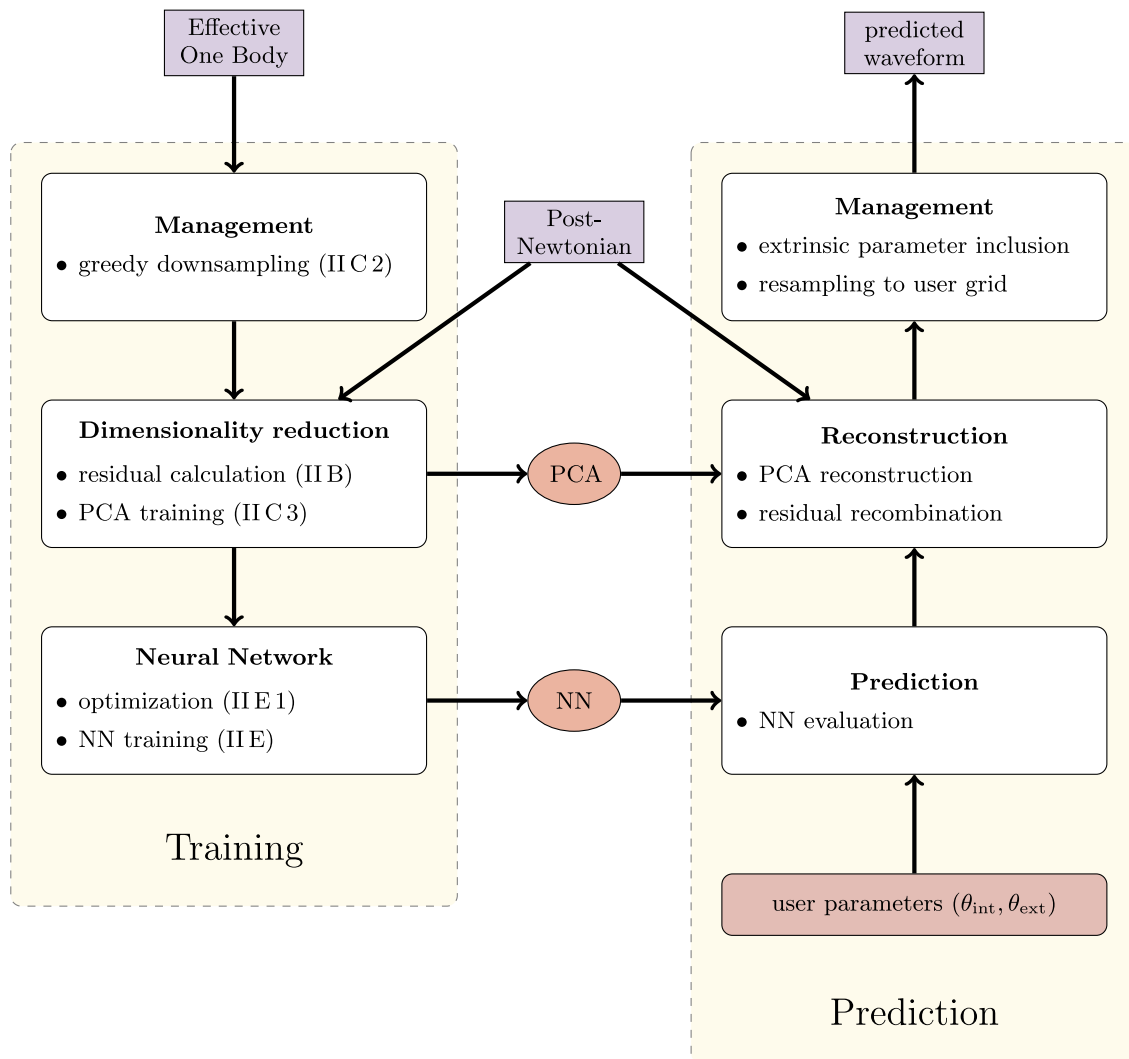
FIG. 1.   Flowchart for the operation of `mlgw_bns`.

This means that waveforms returned by `mlgw_bns` are aligned with the corresponding PN ones, as opposed to the EOB ones. Since the prediction of the merger time within `mlgw_bns` is modeled on the EOB one, this means that the predicted waveforms' mergers fluctuate by the same few tens of milliseconds. This is inessential for the purposes of inspiral-only parameter estimation, but it can be problematic if we wish to extend the inspiral model with one for the postmerger [33,58]. A solution to this could be to reconstruct the time-shift dependence on the parameters $\Delta t(\theta_{\text{int}})$, and de-shift the predicted waveforms after generating them with the PN alignment; this is however not implemented in version `0.12.0` of `mlgw_bns` used in this work.

C. Dimensionality reduction

Neural networks can be small and simple if the dimensionality of the data they must operate on is itself small. Fortunately, the default representation of residuals (or waveforms) in frequency space contains a large amount of redundancy; this section discusses our approach to reducing the dimensionality of its representation.

There are three steps employed within `mlgw_bns` to this end: two of them are different techniques of decreasing the number of points in frequency space the residuals are sampled at, and the third is PCA.

The orders of magnitude for how many floating point numbers are needed to represent waveforms or residuals starting at 5 Hz after each of these steps are as follows (see also Figs. 3 and 4 for a breakdown of where these points are used in frequency space):

(1) the default uniform frequency spacing requires $\sim 2 \times 10^7$ points per waveform, scaling with $f_0^{-8/3}$;

(2) the multibanding approach reduces this to $\sim 2 \times 10^5$, scaling with $f_0^{-5/3}$ [59];

(3) the dataset-dependent greedy downsampling approach reduces this to $\sim 3 \times 10^3$ for full waveforms or $\sim 10^3$ for residuals; and
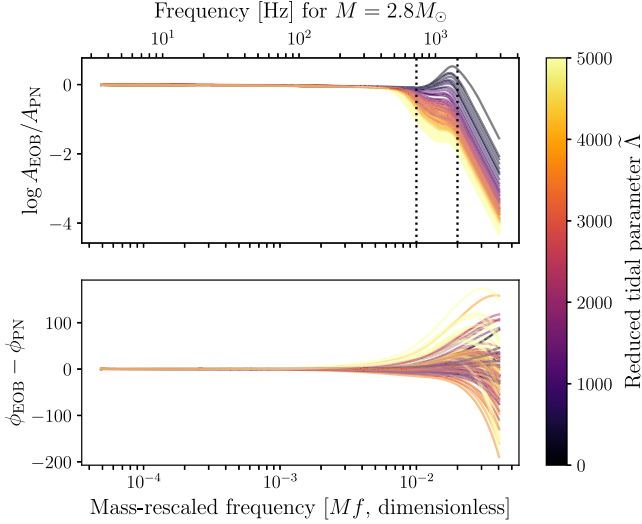
FIG. 2. Residuals of 100 EOB waveforms to their PN counterparts. The EOB waveforms are chosen according to a uniform distribution in parameter space.
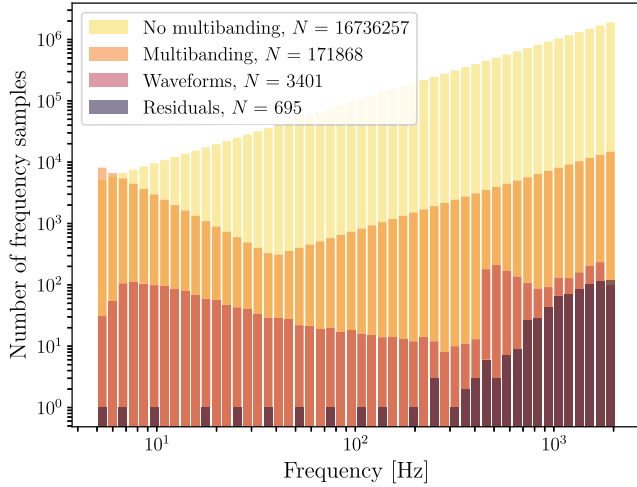

FIG. 3. Comparison of various ways to sample the amplitudes of a waveform. We show histograms of the arrays of frequencies used for the sampling, in the cases of no multibanding (uniform spacing $\Delta f = $ const), multibanding (discussed in Sec. II C 1), and training the greedy algorithm discussed in Sec. II C 2 on 128 waveforms or 128 sets of residuals, computed as discussed in Sec. II B.

(4) the PCA representation, finally, only requires $\sim 3 \times 10^1$ numbers per waveform.

Uniform spacing is never used within `mlgw_bns`; a small number of waveforms is generated directly with the multibanded grid in order to train the greedy downsampling, and once this is done all further waveforms are generated on the smaller greedy downsampling grid. This means that, even when starting from a very low initial frequency, we can easily work with a dataset of waveforms within the RAM of a laptop.
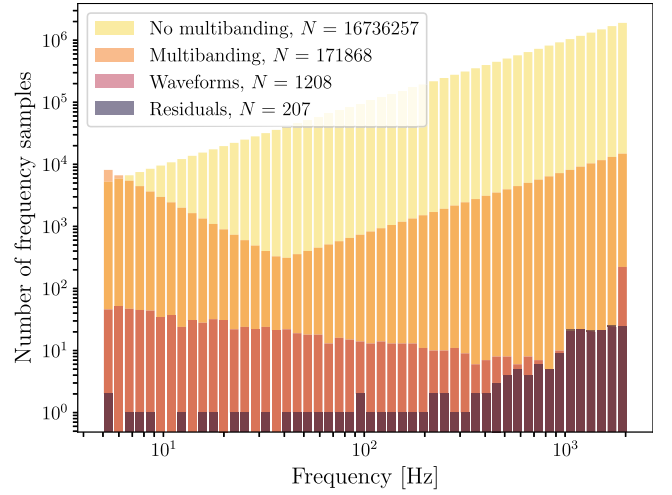

FIG. 4. Same as Fig. 3, but training the greedy algorithm to reconstruct the phase of the same waveforms.

1. Multibanding

"Multibanding" is the name we give to a technique of generating a frequency grid which is much smaller than the uniform one, not very dependent on the specifics of the dataset, and which may still be used to get a good representation of compact binary coalescence (CBC) waveforms.

The starting point is the observation that a CBC signal will always have a specific chirping profile, with high-frequency information only contained in a short (in time) section at the end. The default frequency array used in signal processing, for a real-valued signal with duration T and time spacing Δt, will be a uniform array from $f = 0$ to $f = 1/2\Delta t$ (the Nyquist frequency), with spacing $\Delta f = 1/T$. As expected, this means there is no information loss; $T/\Delta t$ real numbers are mapped to $(2\Delta t)^{-1}/T^{-1}$ complex numbers. This array describes high- and low-frequency information for all times; in the CBC case this entails a lot of redundancy, since it is already known ahead of time that for the overwhelming majority of the signal there will be no high-frequency information. We may construct a frequency array which is "aware" of this behavior [59,60]. We start from the fact that the duration of a CBC signal starting from a frequency f_0 is $T \propto f_0^{-8/3}$, with a proportionality constant that can be analytically derived at Newtonian (0PN) order and which depends on the mass and the mass ratio [61]:

$$T = \frac{5}{256}(\pi f_0)^{-8/3} M^{-5/3}/\nu. \qquad (8)$$

Then, we can make a frequency array for which the frequency spacing at each frequency is $\Delta f(f) \approx 1/T(f)$. This will mean we sample the low-frequency region much more finely than the high-frequency one, but locally each frequency band is described with the correct level of detail.

The approach used within `mlgw_bns` differs from the one used by Ref. [59] in two aspects. First, whereas they approximate the uniformly-varying Δf by dividing the frequency domain into bands and using a different, uniform frequency for each of those, we construct a frequency array with continuously-varying spacing. Second, while they extend this sampling into the high-frequency regime, we use it only for frequencies lower than a certain pivot, typically $f_{\text{pivot}} \approx 40$ Hz, while for higher frequencies we use uniform sampling. This is a conservative choice, motivated by the fact that at high frequency the 0PN expression for the time to merger cannot be expected to hold, combined with the fact that a uniform array with the spacing defined by $\Delta f = 1/T(40 \text{ Hz}) \approx 0.02$ Hz is not a large computational burden, resulting in only a few tens of thousands of points. This approach, that we call *multi-banding*, needs to know something about the dataset: while the mass is kept fixed during the training, the mass ratio cannot be. The dependence is $\Delta f \propto 1/T \propto \nu$, and ν scales inversely with the mass ratio q (which is >1 here). Therefore, the *smallest* Δf we should use as a lower bound corresponds to the largest q within the dataset; note, however, that this characteristic is shared by the uniform sampling, which is also defined by the three quantities $f_{\min}$, $f_{\max}$, and Δf.

Figures 3 and 4 show histograms for the multibanding approach compared to the standard, uniform-in-frequency approach, for the case of waveforms starting at 5 Hz. The uniform-in-frequency grid looks tilted in the histogram since the bins represent logarithmic frequency intervals, which increase in absolute width (Δf) as the frequency increases. The general pattern to observe is that, as we make more and more assumptions about the waveforms we need to represent, the frequency array can shrink. Multibanding is a rather safe choice, since it makes no more assumptions than uniform sampling, but it still provides at least an order-of-magnitude improvement in typical cases. The lower two histograms, labeled "Waveforms" and "Residuals," show the numbers of points that can be achieved when greedily selecting frequencies by requiring that they allow us to reconstruct full EOB waveforms or their residuals (described in Sec. II B) respectively.

2. Downsampling

While the multibanding reduces the size of the frequency arrays by orders of magnitude, especially for very low initial frequencies, we can do even better if we allow a heavier dependence on the specific dataset. Specifically, we can determine a set of points in frequency space such that any waveform in the dataset, if given at those points only, can be interpolated and retrieved at all frequencies within a certain accuracy.

In order to achieve this goal, a greedy optimization technique is used. First, a set of waveforms is generated on the grid described in the previous section. These waveforms are then downsampled to a sparse grid, which can initially just consist of the endpoints of the domain, and resampled with a cubic spline.[3] The reconstruction error can then be measured for each of these waveforms; new points are added to the grid where it is worst. This procedure is iterated until all the given waveforms can be reconstructed within a certain tolerance, which we select to be 10^{-5} for both amplitude and phase. The downsampling is performed separately for amplitude and phase.

As the diagram in Fig. 1 shows, when reconstructing a waveform the "residual recombination" step happens before the "resampling to user grid" step. This means that this downsampling procedure, which by itself is a generic algorithm, is applied to the *full EOB waveforms* as opposed to the residuals described in Sec. II B.

While this requires us to use a slightly larger frequency grid (but still with $<10^4$ points), it was found to be generally faster than the alternative.

3. Principal Component Analysis

Once the waveform has been downsampled, it is represented with n_A numbers for the amplitude and n_ϕ for the phase. Its dimensionality can be further reduced using PCA. We collect all the residuals corresponding to each waveform in an array $x = [\Delta A, \Delta \phi] \in \mathbb{R}^{n_A + n_\phi}$ and construct a training dataset out of such arrays, $\{x_i\}_i$, of which we may compute the mean $\mu = \langle x \rangle$ and the covariance matrix

$$C = \langle (x - \mu)(x - \mu)^\top \rangle. \tag{9}$$

This (symmetric, positive definite) matrix is diagonalized as $C = VDV^\top$, where $D = \text{diag}(\lambda_i)$ is a diagonal matrix containing the eigenvalues of the covariance matrix, ordered so that $\lambda_i \geq \lambda_{i+1}$. The columns of V are the eigenvectors and, because of the ordering, the first k eigenvectors correspond to the k largest eigenvalues. Projecting a vector x onto the span of these k eigenvectors allows us to approximately represent it with only k numbers. Specifically, if U is the $(n_A + n_\phi) \times k$ submatrix of V consisting of the k eigenvectors corresponding to the largest eigenvalues of the covariance matrix C, we explicitly write the forwards and backwards transformations for x into its low-dimensional representation $\tilde{x}$:

$$x \to \tilde{x} = U^\top (x - \mu) \tag{10a}$$

$$\tilde{x} \to x = U\tilde{x} + \mu. \tag{10b}$$

---

[3]Cubic interpolation was found to be a good middle ground when accounting for computational complexity (which increases with interpolation order) and greedy grid size (which decreases with interpolation order).

The number of principal components to keep can be tuned depending on the required final fidelity, including more of them increases the evaluation time of each waveform.

For simplicity, for the remainder of this work we always retain 30 principal components. In principle this number could also be tuned, and its current value was mainly chosen to be "safely large." This is confirmed by Fig. 6: the reconstruction fidelity grows in a roughly linear fashion with the number of training points and its accuracy is never hampered by the number of PCA up to fidelities $\bar{\mathcal{F}} \lesssim 10^{-5}$. As we will discuss in Sec. IV B, even with this possibly suboptimal value of k our model is fast enough not to be the bottleneck in the evaluation of the likelihood.

D. Frequency band

As our detectors improve their sensitivity at low frequency, it is crucial to have a model which can be conveniently evaluated there. In this section, we discuss the frequency band in which our model is trained, and how we may overcome the inherent limitation of only training down to a given frequency.

For the default model, which is provided with version `0.12.0` of `mlgw_bns` and whose performance is discussed in this work, the frequency range for which validity is guaranteed is [5, 2048] Hz, while the range of valid total masses is $[2, 4]M_\odot$; this means, as we shall discuss below, that the reference-mass model is trained in the range $\approx [3.57, 2926]$ Hz.

When the user requests frequencies within the training range, the model is able to directly yield a prediction; however this may be limiting, especially when considering multiband observations. The waveform at frequencies lower than the ones in the training range is well described by the PN approximation; therefore, waveforms predicted by `mlgw_bns` are natively hybridized with PN ones at low frequency, as Sec. II D 3 below describes.

1. Mass rescaling

As mentioned in the introduction, we exclude the total mass M from the training parameters since the waveform only depends on the combination Mf; this affects the frequency band in which we must train our model.

Suppose the user requires a waveform $h(f; M, \theta)$ with total mass M. Then, the overall waveform is computed within `mlgw_bns` as

$$h(f) = \frac{M}{M_{\text{ref}}} h\left(\frac{fM}{M_{\text{ref}}}; M_{\text{ref}}, \theta\right), \qquad (11)$$

which means that the user-given frequency grid will be shifted by a factor M/M_{ref}.

In order for this to yield a valid waveform, however, the shifted frequencies must still lie within the model's training frequency range.

Therefore, if we want our model to be applicable for all frequencies in a range $[f_1, f_2]$ and for all masses in a range $[M_1, M_2]$ we need to train the reference-mass model in a range

$$f \in \left[f_1 \frac{M_1}{M_{\text{ref}}}, f_2 \frac{M_2}{M_{\text{ref}}}\right]. \qquad (12)$$

2. High frequency bound

The model we are training on, `TEOBResumSPA`, describes the inspiral up to merger, which in the mass range of interest typically happens above 2 kHz. After the merger, the remnant (a short- or long-lived neutron star, or a black hole) will emit a postmerger GW signal for which models exist [32,33,62–65], but which is considered separately from the EOB waveform; after the merger frequency, `TEOBResumSPA` waveforms are tapered with a powerlaw in the amplitude, $A_{\text{EOB}} \propto f^{-10/3}$, and a linear relation in phase, $\dot{\phi}_{\text{EOB}} = \dot{\phi}(f_{\text{max}})$, for $f > f_{\text{max}}$ [[37], Eqs. (S11) and (S12)]. This scaling is enforced as to ensure that the inverse Fourier transform of these waveforms is close to the time-domain waveform. Also, it means that the amplitude is guaranteed to remain positive (albeit quickly diminishing) at high frequency.

However, this implies an issue in the residuals computation of Eq. (6): the baseline PN approximant is written as a power series in $v = (\pi M f)^{1/3}$, which means that there is no guarantee that A_{PN} will remain positive in the high-frequency regime, and indeed in practice, it often does become negative, which means that our residuals defined in Eq. (6) diverge.

We fix this by choosing a maximum frequency for the validity of the PN model, and setting its amplitude to a constant value after that. This is not done "sharply," since that would propagate a discontinuity to the prediction; instead, we smoothly connect the expressions within an interval $[f_1, f_2] = [0.01/M, 0.02/M]$ as follows: for all $f \in [f_1, f_2]$ we write

$$A_{\text{PN}}^{\text{new}}(f) = (1 - \zeta(x(f)))A_{\text{PN}}(f) + \zeta(x(f))C, \qquad (13)$$

where $\zeta: [0, 1] \to [0, 1]$ is chosen so its derivative at the boundaries vanishes; specifically, we use

$$\zeta(x) = \frac{1}{2}(1 - \cos(\pi x)), \qquad (14)$$

while

$$x(f) = \frac{f - f_1}{f_2 - f_1}. \qquad (15)$$

The constant C is chosen to be equal to 20 in natural units; this is somewhat arbitrary, but it is roughly the value

attained by A_{EOB} at $f \sim 0.02/M$, as demonstrated by the first panel in Fig. 2: the value at $Mf = 0.02$ is $\log A_{\text{EOB}}(Mf = 0.02)/C$, and one can see that it changes sign as we vary $\tilde{\Lambda}$.

This shows that $C = 20$ is a reasonable middle ground for this parameter. This choice will only have an impact on the network's ability to learn the residuals; if they are reconstructed correctly and the same modified PN model is used both in training and reconstruction, the specifics of the modification do not matter, and the high-frequency continuation of our waveforms is equal to the EOB one described at the beginning of this section.

For simplicity, for all frequencies higher than the maximum training one, we return a waveform which is identically equal to zero.

3. Low-frequency bound

For a typical BNS, a frequency of 5 Hz corresponds to about two hours before merger. This is close to the lower frequency limit for a ground-based detector, but for a multiband observational campaign (including space- or Moon-based detectors) having a model able to be evaluated at arbitrarily low frequencies is very convenient.

The architecture in `mlgw_bns` makes this easily achievable: since we are reconstructing residuals from a PN baseline, we may evaluate the waveform at arbitrarily low frequencies by setting the residuals to zero and just yielding the PN waveform, which below 5 Hz is a very good approximation of the true waveform; as Fig. 2 shows, the residuals approach zero in the low-frequency regime.

For the phases, by subtracting an arbitrary linear term we can achieve $\phi(f_{\min}) = 0$ exactly, and $\dot{\phi}(f_{\min}) \approx 0$ to quite good accuracy, therefore we can simply yield PN phases below $f_{\min}$ and our prediction above it. For the amplitudes, this is not the case, and a discrepancy of the order of $\Delta \log A \sim 5 \times 10^{-3}$ remains.

This discontinuity is fixed by a smoothing procedure: PN amplitudes corresponding to frequencies between $f_{\min}/2$ and $f_{\min}$ are rescaled, so that the output of the model is

$$A(f) = \begin{cases} A_{\text{PN}}(f) & f < f_{\min}/2 \\ A_{\text{PN}}(f) + \Delta A \zeta\left(\frac{2f}{f_{\min}} - 1\right) & f_{\min}/2 \le f \le f_{\min} \\ A_{\text{EOB}}(f) & f \ge f_{\min}. \end{cases}$$

$$(16)$$

E. Neural network

A feed-forward neural network is trained to reconstruct the map $\theta \to \tilde{x}$, where θ is the vector of the five intrinsic parameters considered, while $\tilde{x}$ is a 30-dimensional PCA representation of the residuals corresponding to the waveform generated by the EOB model with the given parameters.

As our neural network we employ a `MLPRegressor` from the SCIKIT-LEARN library [66], and the training is performed with the Adam algorithm for stochastic gradient descent [67].

As it is common, the parameters θ are rescaled to have mean 0 and standard deviation 1. After the PCA reduction, each component in the vector $\tilde{x}$ natively has comparable variance, but we may arbitrarily rescale them, which is equivalent to rescaling the eigenvectors in the matrix U defined in Sec. II C 3. Also, we know that the eigenvectors corresponding to the largest eigenvalues λ_i "matter more," in that they explain more variance. Therefore, as a preprocessing step we introduce a fixed rescaling of the vector $\tilde{x}$, as $\tilde{x}_i \to \tilde{x}_i \lambda_i^\alpha$ for some tunable choice of $\alpha \ge 0$. The distance used during the training is then simply the Euclidean one between these rescaled $\tilde{x}$.

1. Hyperparameter optimization

Several hyperparameters, which determine the network's properties and performance, must be chosen before training, such as the number and size of hidden layers in the network, the activation function, the conditions for the termination of the training, the coefficient for the regularization term, and the coefficient α defined above. For a complete list, see Appendix B, which details all the hyperparameters used within the default network discussed here.

The optimal set of hyperparameters may vary as the number of training waveforms used to train the network may change. Heuristically, we might imagine that a complex network with many layers would be the best choice with many thousands of training waveforms, while it would overfit when using only a hundred waveforms for the training, for which the optimal configuration would be a smaller network. The specific dependence of the reconstruction efficiency on these parameters is, however, high dimensional and hard to explore since evaluating each point requires us to train the whole network.

We evaluate each possible set of hyperparameters by computing its average reconstruction error on a validation dataset, generated independently but from the same distribution as the training dataset; the reconstruction error is measured as the distance defined by

$$\text{dist}^2(\tilde{x}_{\text{orig}}, \tilde{x}_{\text{pred}}) = \frac{\|x_{\text{orig}} - x_{\text{pred}}\|^2}{n_\phi + n_A}, \qquad (17)$$

where $n_\phi + n_A$ is the dimensionality of the vector x, as defined in Sec. II C 3; the distance is written in terms of the vectors $x = [\Delta A, \Delta \phi]$, reconstructed from the PCA-reduced $\tilde{x}$ predicted by the network.

The hyperparameters are optimized with the Optuna package [68] using a multiobjective tree-structured Parzen estimator [69], where the two cost functions being simultaneously optimized are

(1) the average reconstruction accuracy on a validation dataset measured as in Eq. (17); and

(2) the estimated time required for the generation of the training waveforms, quantified by 100 ms times the number of training waveforms, plus the time needed to train the network.

The training and validation datasets are randomized in each iteration.

The use of these two "opposed" cost functions allows for a Pareto front of optimal parameters to be computed. This is a collection of parameter sets corresponding to different training dataset sizes; once this optimization has been run, for any given dataset size we have a set of good hyper-parameters to train the network.

Such a collection—with dataset sizes ranging from 50 to 10^5 training waveforms—is provided with version 0.12.0 of mlgw_bns, and Fig. 5 shows the validation errors as a function of training dataset size. When creating a new model, a lookup may then be performed to recover the locally optimal hyperparameters for the amount of data available to the model. This is efficient since it allows us to train new networks without rerunning the optimization when the parameter space utilized remains relatively similar to the one used during the optimization procedure; we have however found that with significant changes to the parameter space (e.g., including versus not including spin) the optimization had to be rerun since it was giving suboptimal results.

III. MODEL PERFORMANCE

A. Accuracy

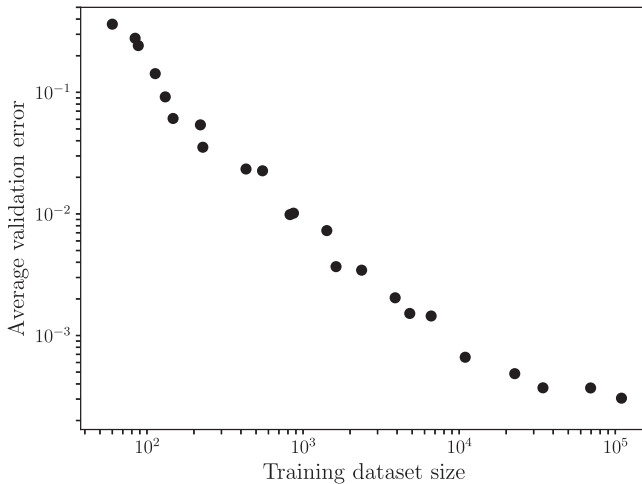Figure 6 shows the mismatches between the recon-structed waveforms and the corresponding EOB ones.

FIG. 5. Pareto front for the hyperparameter optimization. The vertical axis shows the average error, computed as in Eq. (17). The flattening observed at large training dataset sizes is not necessarily real; computational constraints prevented a large amount of trials to be performed in that region.

The mismatches are computed on validation datasets generated with the same distribution as the training ones, but with differently-seeded random number generators. The mismatches are computed according to the predicted Einstein Telescope PSD, ET-D [53,54]. As shown by the figure, the accuracy measured through the mismatch $\bar{\mathcal{F}}$ exhibits a roughly linear behavior $\bar{\mathcal{F}} \sim 1/N_{\text{train}}$.

The reconstructed residuals corresponding to the best model of Fig. 6 (trained with $2^{17} = 131072$ waveforms) are shown in Fig. 7. As one might expect, the residuals significantly differ from zero only in the high-frequency region, like the original residuals. When considering the magnitude of the phase residuals, note that the logarithmic frequency axis distorts what may be linear trends: the temporal alignment chosen in the plot was not optimized to correspond to the best-match one, but instead to align the waveforms at low frequency.

B. Speed

The evaluation times for mlgw_bns are shown in Fig. 8 and compared to the evaluation times of TEOBResumSPA. The significant acceleration provided by mlgw_bns is maximized when using compressed frequency grids with a small number of points, where it attains speedup factors
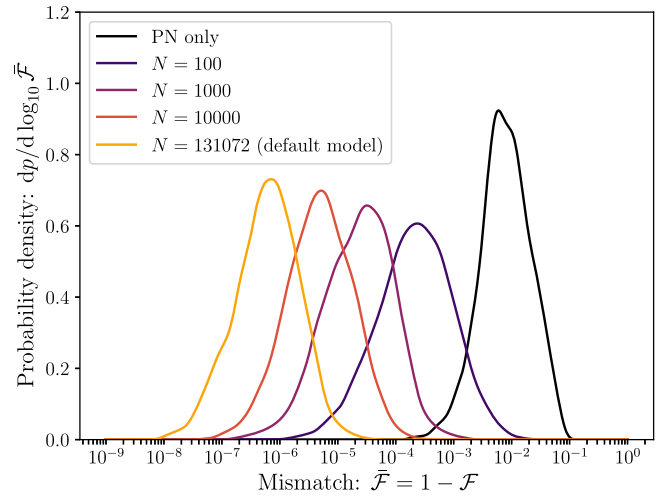

FIG. 6. Kernel Density Estimate representation of the mis-matches between the waveforms reconstructed by mlgw_bns and the corresponding ones generated by the reference waveform generator, TEOBResumSPA, for uniformly-distributed sets of parameters θ_{int} in the training ranges, and with constant total mass $M = M_{\text{ref}} = 2.8 M_{\odot}$. The curve labeled as "PN only" is obtained by comparing the baseline PN waveforms with the corresponding EOB ones, i.e., setting the reconstructed residuals to zero; for the other curves we use the number indicated for both the training of the PCA and for the training of the network, so the overall number of waveforms used is twice N. The same 4096 validation waveforms are used to generate each curve. The mismatch is computed according to the predicted Einstein Telescope PSD, ET-D [53,54], within the band [3.57, 2926] Hz (see Sec. II D).
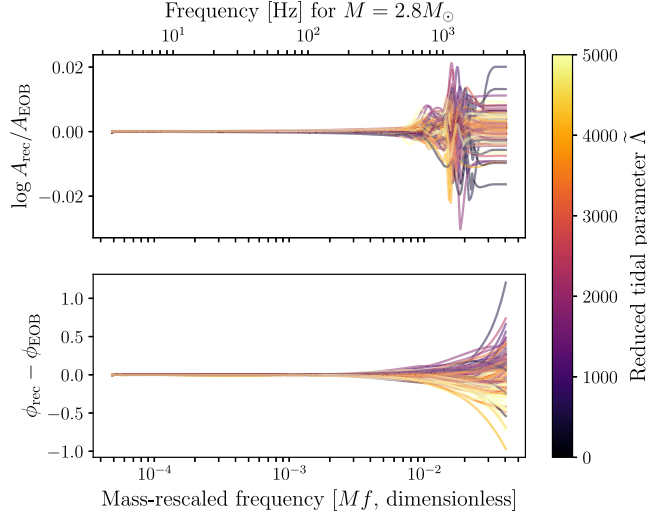
FIG. 7. Residuals of 100 reconstructed waveforms to the reference EOB ones. The parameters for them are uniformly distributed.
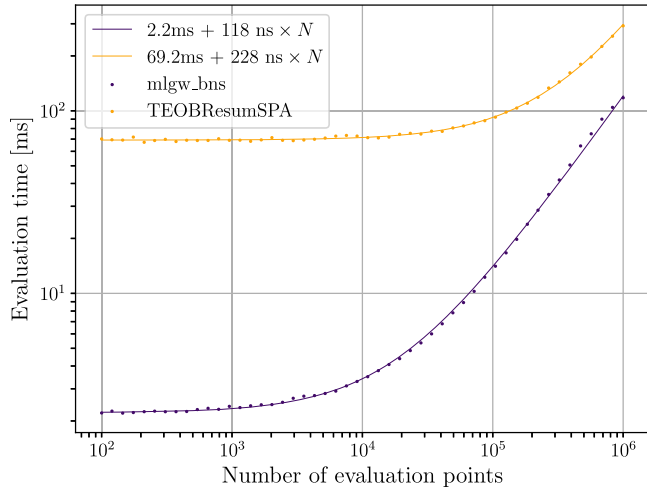

FIG. 8. Benchmarks of the evaluation time required for one waveform, with `TEOBResumSPA` and with `mlgw_bns`. Details on the benchmarking procedure can be found in Appendix A. For both approximants, we also show a fit with a model $t = t_o + t_p N$.

of ~35. The high template efficiency allows `mlgw_bns` to outperform not only `TEOBResumSPA` in waveform generation, but also all other state-of-the-art EOB surrogate models available, as demonstrated in Appendix A.

Both templates exhibit a similar behavior in the number of sampling points: $t(N_{\text{sample}}) \sim t_o + t_p N_{\text{sample}}$. There is an approximately constant cost to evaluate the waveforms at small values of N_{sample}, while for large N_{sample} the evaluation time scales linearly. This is due to the fact that, for both templates, there are operations that are approximately independent on the number of evaluation points. For `mlgw_bns`, these are running the parameters through the neural network and recomposing the result through PCA. For `TEOBResumSPA`, the solution of the

Hamiltonian flow using the postadiabatic EOB iteration (at fixed number of points) and the subsequent ODE evolution for the last few orbits before merger [34]. The linear regime is instead, for both templates, caused by the time to interpolate the waveform to each of the finely-spaced user-given frequency points, and performing other linear-time operations such as combining amplitude and phase into the Cartesian representation of the waveform.

The linear-time operations taken by the two approximants are comparable; `TEOBResumSPA` is implemented in C and `mlgw_bns` in Python, but several components in the latter are just in time compiled thanks to `numba` [70]. While the constant t_p might be whittled down by optimizing the implementation, the linear term cannot be completely removed—the program will have to do at least a few floating point operations for each point we are resampling at. Therefore, if we want fast waveform evaluation it is important to use as small a number of points as we can, while retaining the desired accuracy.

Several approaches have been suggested towards this goal for PE purposes: the simpler ones are similar in spirit to what has been discussed in Sec. II C 1, using a smart coarser sampling than what the "natural" FFT grid would be. More sophisticated approaches can be ROQs (discussed below in the context of PE) or relative binning [71,72].

In Tab. I we show a breakdown of the use of time within an evaluation of `mlgw_bns`, in the case of 1000 grid points.

IV. PARAMETER ESTIMATION

To showcase the benefits brought by our model in a realistic setting, we perform PE studies on the binary neutron star (BNS) transient GW170817 [48,49]. In Sec. IV A, we first perform a full-scale validation, showing the compatible results of GW inference using `mlgw_bns`, compared to the ones obtained with `TEOBResumSPA`.

TABLE I. Timing breakdown for the evaluation of a waveform on 1000 grid points with `mlgw_bns`. Values will fluctuate across evaluations, this table is only meant to be indicative of the ratios between them.

Task	Time [μs]	Subtask	Time [μs]
Resampling	841	Spline creation	728
		Spline evaluation	113
PN evaluation	653	Amplitude	434
		Phase	219
PCA + NN	397	NN	326
		PCA	41
		Miscellaneous	30
Postprocessing	289	Include extrinsic	157
		Compute $h = Ae^{-i\phi}$	40
		Miscellaneous	90
Total	2180		

Then, in Sec. IV B we discuss and apply compression techniques capable of reducing the number of frequency nodes on which `mlgw_bns` needs to be evaluated for PE purposes. This step allows us to fully exploit the benefits of our model, which displays the largest gain compared to `TEOBResumSPA` for a smaller number of frequency nodes (see Fig. 8). Section IV C finally repeats the PE analysis combining `mlgw_bns` and such compression methods, showcasing more than order of magnitude speed gain obtainable with our ML technique against `TEOBResumSPA` in a full-fledged PE analysis.

In particular, we analyze the (deglitched) GWOSC data of LIGO and Virgo centered around GPS time 1187008857.6 with a sampling rate of 4096 Hz and a duration of 128 s, considering the frequency range from [23, 2000] Hz. Our PE relies on the MPI-parallelized Bajes pipeline [38] and the DYNESTY [73] nested sampler. The reported errors correspond to the 90% confidence intervals and the log symbol refers to the natural logarithm. The mass prior is chosen to be flat in the mass components $m_{1,2}$, although the sampling is then performed in $(\mathcal{M}, q)$, with

ranges wide enough to capture the full posterior width. We sample on aligned-spin components, with an isotropic prior bounded by $\chi_{1,2} \leq 0.5$. The prior on the tidal parameters is uniform in the ranges $\Lambda_{1,2} \in [5, 5000]$ and the luminosity distance employs a volumetric prior in $D_L \in [1, 75]$ Mpc. Other priors are set according to standard prescriptions in GW astronomy [38]. We do not assume prior knowledge on electromagnetic counterparts. We include spectral calibration envelopes with ten logarithmic-spaced nodes for each detector. For an overview of Bayesian inference of GW signals see Refs. [2,38,74,75].

A. Full grid `mlgw_bns`—`TEOBResumSPA` comparison

Using the settings discussed above, GW170817 is analyzed with `mlgw_bns` and `TEOBResumSPA` in order to compare performances and verify the consistency of the results. The sampling employs 3000 live points, an evidence tolerance of 0.1, a maximum number of Markov-Chain Monte Carlo steps of 12000 and five auto-correlation times before accepting a point. We analytically marginalize over the coalescence time t_c and phase ϕ_c.
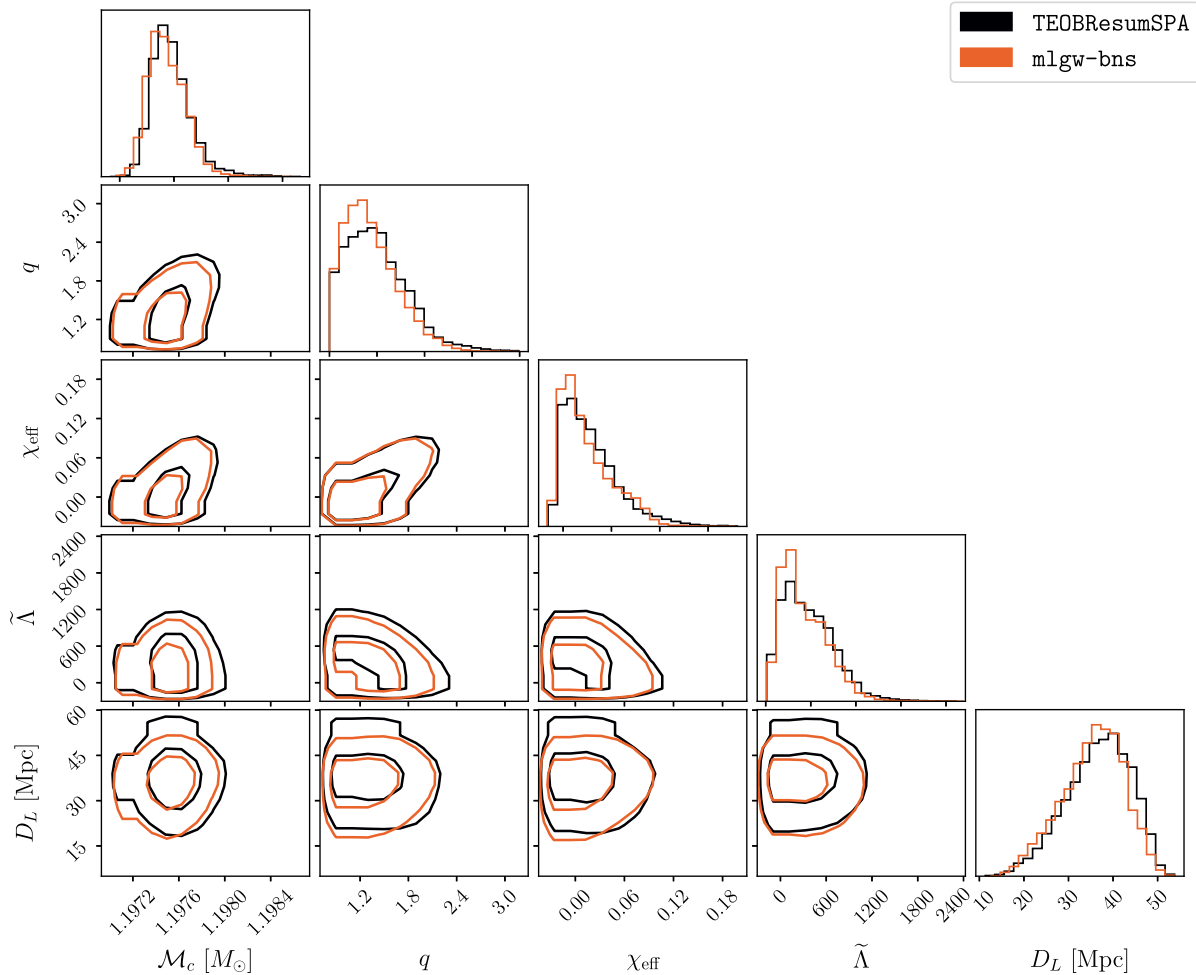

FIG. 9.   Corner plot the posterior distribution for selected parameters reconstructed for GW170817, with `mlgw_bns` (orange) and `TEOBResumSPA` (black). The contours report the 50% and the 90% credibility regions.

The two waveform approximants achieve compatible measurements, within the stochasticity of the sampler. Figure 9 shows the comparison between `mlgw_bns` and `TEOBResumSPA` posterior distributions for selected parameters of interest.

We recover $\mathcal{M} = 1.1975^{+0.0003}_{-0.0002} M_{\odot}$, the mass ratio is constrained to $q < 2.07$ at the 90% confidence level and the reduced tidal parameter corresponds to $\tilde{\Lambda} = 365^{+522}_{-254}$. The recovered posteriors are consistent with previous similar studies [37,38,48,49,76]. Moreover, the two models recovered similar Bayes' factors ($\log \mathcal{B} \simeq 482$), and signal-to-noise ratios (SNR = 32), validating the faithfulness of `mlgw_bns` with respect to the training template in a realistic application. As an additional check, in Fig. 10 we show the difference in log likelihood computed at all the points in both posteriors is compatible with zero, with fluctuations of order $\Delta \log \mathcal{L} \sim 0.1$; the small apparent discrepancies between the posteriors are due to sampler convergence.

We observe only a mild improvement in execution time for `mlgw_bns` compared to `TEOBResumSPA`. This is expected given the uniform frequency grid with $(f_{\max} - f_{\min}) \times T = (2000 - 23) \times 128 = 253056$ evaluation points. In fact, Fig. 8 shows that for this number of points the advantage in generating waveforms using `mlgw_bns` is not enormous. Significant speedups can instead be achieved by relying on grids smaller than 10^4 points. This naturally calls for the usage of compression techniques, capable of restricting the required number of frequency nodes used in computing the likelihood, the subject of the remainder of this section.
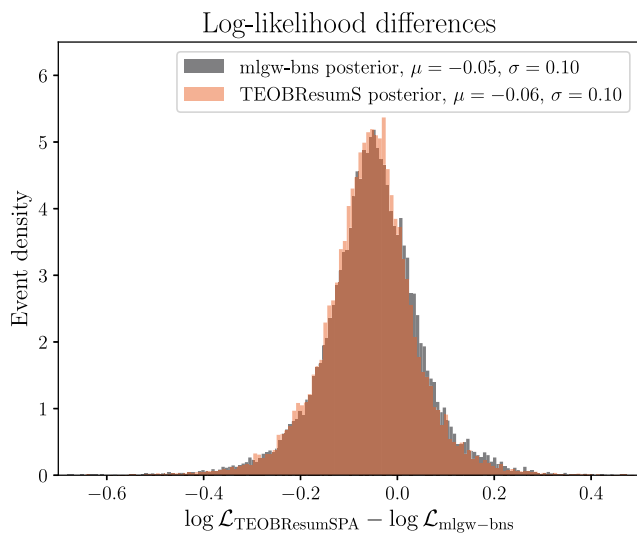

FIG. 10. Histogram of the difference in GW170817 likelihoods evaluated on all points of the `TEOBResumSPA` and `mlgw_bns` posteriors shown in Fig. 9. For each point in the posterior distributions, the likelihoods are evaluated and compared; the time and phase shifts are marginalized analytically, therefore we maximize them for each likelihood independently (finding slightly different timing, for the reasons discussed in Sec. II B).

B. Reduced order quadrature construction

Reduced order modeling, which is referred to as ROQs in GW astronomy when combined with discrete empirical interpolation techniques, is a method of eliminating information redundancy present in sets of parametric functions (in our case, the gravitational waveforms as functions of the physical parameters of the binary system, such as masses and spins) when evaluated on a discrete set of points (in our case, the frequency grid). By selecting a small number of waveforms' "basis elements" and an equal number of discrete interpolation frequency points, ROQs are capable of dramatically speeding up both waveform evaluation and integrals involving them, such as the Wiener inner products [see Eq. (4)] entering the standard GW likelihood. This is achieved by sufficiently accurate—and fast to evaluate—*interpolants*, built on a large training dataset. In the context of GW astronomy, early development and applications of ROQs to GW searches were presented in Refs. [77,78]. An extended mathematical analysis (notably, including convergence estimates) was presented in [79], while the construction of surrogate models using related techniques was pioneered in Ref. [39]. Applications to PE were introduced in [80,81], and the extension to precessing signals PE was achieved in [60], also including many improvements such as mass-frequency partitions and an adaptive frequency sampling strategy. Reduced order quadrature acceleration of tests of GR was considered in [82]. Most of the methods used in the aforementioned applications are implemented in the GreedyCpp code.[4] More recently, ROQs of precessing signals containing higher harmonics were presented in Ref. [83], while Ref. [5] used ROQ methods to demonstrate the feasibility of analyzing BNS merger signals detected by the next generation of ground-based detectors. The interested reader may refer to [39,60,79] for an introduction to the concepts used below.

Reference [83] introduced a set of modifications in how the initial basis elements are constructed compared to previous literature, aiming at improving the efficiency of basis construction. The related algorithm was released in a public Python package, labeled PyROQ.[5] We modified and generalized this algorithm, added numerical stability checks, and restructured the software to make it more modular and easily usable with modern (typically Python-based) waveform approximants. Details of our algorithm, labeled `JenpyROQ`,[6] and GW170817 ROQ interpolants construction are presented in Appendix C. For the PE analysis discussed above, we obtained a sufficiently accurate basis with 267 (10) linear (quadratic) basis elements, achieving a linear (quadratic) frequency axis reduction factor of 950 (25300).

---

[4]Available at: bitbucket.org/sfield83/greedycpp.
[5]Available at: github.com/qihongcat/PyROQ.
[6]Available at: github.com/GCArullo/JenpyROQ.

TABLE II. Waveform generation (t_{wf}) and likelihood inner-products (t_{ip}) timings when using ROQs or a full frequency grid evaluation. We report results for both a single ($N_{det} = 1$) and three detector network ($N_{det} = 3$). The total likelihood evaluation time is simply $t_{tot} \simeq t_{wf} + t_{ip}$, since other likelihood operations costs are comparatively negligible. The ROQ approximation results in a PE speedup factor of 18 (12) in the one (three) detector case.

	Timings [ms]			
(N_{det}, ROQ)	(1, no)	(1, yes)	(3, no)	(3, yes)
t_{wf}	69.8	2.2	69.8	2.2
t_{ip}	15.3	2.5	45.9	7.5
t_{tot}	85.2	4.7	115.7	9.7

C. Parameter estimation with reduced order quadrature interpolation

To predict the expected speedup on a PE run using the ROQ interpolants described above, it is sufficient to compute $t_{tot} = t_{wf} + t_{ip}$, where t_{wf} indicates the waveform (Eq. 2) generation time, while t_{ip} indicates the evaluation time of the

likelihood inner products (including interpolants evaluation); all other operations (e.g., detectors projection) are negligible compared to these two costs. Typical values for these times when using an ROQ technique or a full frequency grid evaluation are reported in Table II. For a single detector, the predicted ROQ speedup factor is 85.2 ms/4.7 ms ∼ 18. For three detectors (the case of interest in our realistic application), the total speedup becomes 115.7 ms/10 ms ∼ 12. These numbers imply that when relying on `mlgw_bns` and an ROQ scheme, the waveform evaluation cost is no longer the dominant one. For this reason, the expected PE speedup (12) is a factor of three smaller than the waveform evaluation speedup (35) inferred from Fig. 8.

We validate this by repeating the GW170817 analysis in the previous section, employing `mlgw_bns` both times but using either a GW likelihood built with the ROQ interpolants constructed above, or a standard likelihood computation. We do not apply time marginalization in this case, since we have not interfaced the ROQ formulation with the time-marginalized likelihood, hence we increase the values of sampler settings to avoid altogether any convergence
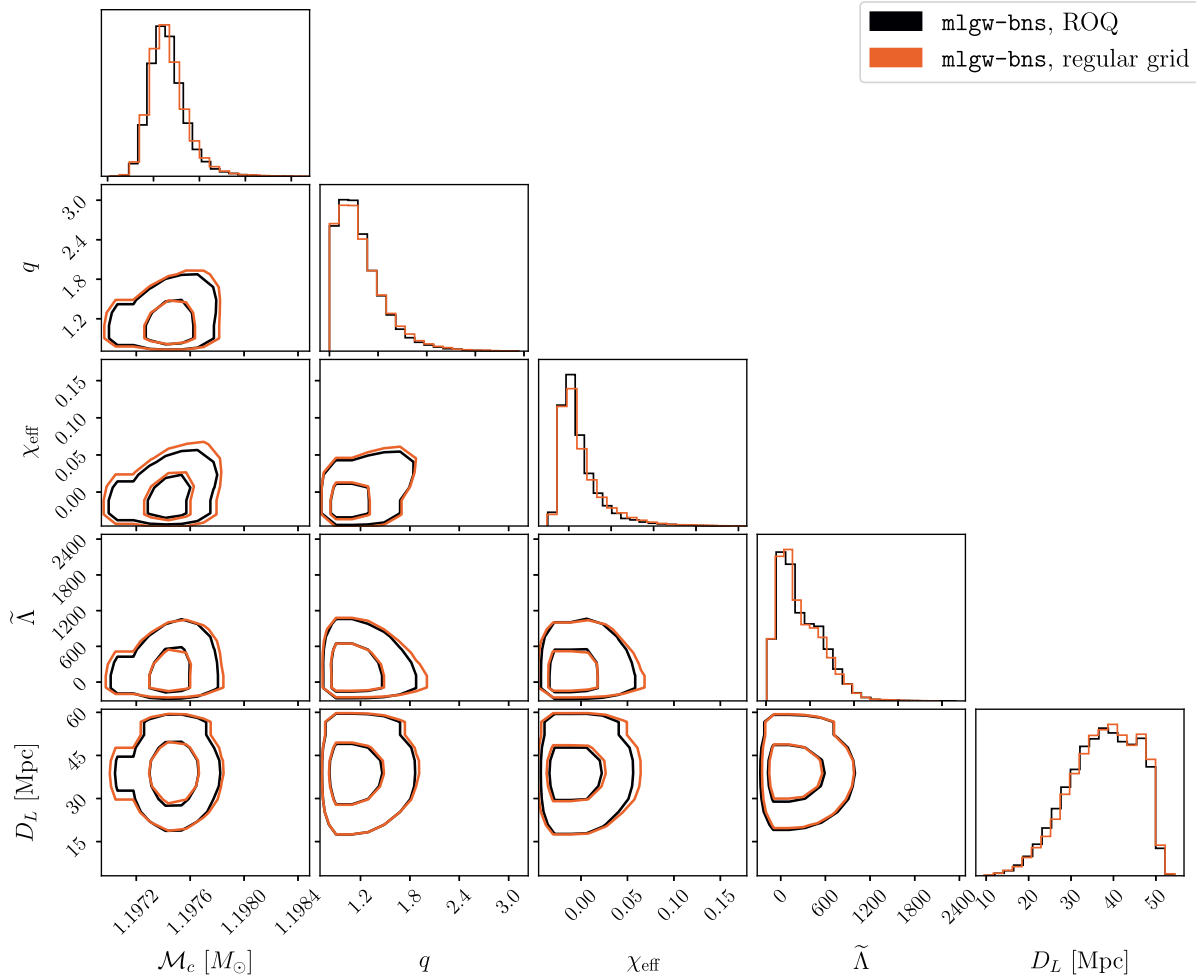

FIG. 11. Corner plot of the posterior distribution for selected parameters reconstructed for GW170817, in both cases with `mlgw_bns`, but when using an ROQ technique or a full frequency grid evaluation. The contours report the 50% and the 90% credibility regions.

issues. We employ 5000 live points, an evidence tolerance of 0.1, a maximum number of Markov-Chain Monte Carlo steps of 12000 and ten auto-correlation times before accepting a point. We explore t_c within the bounds [24.7, 25.0] s, using a discretization composed of 3000 points. Parameter estimation results obtained with the ROQ settings discussed above or with the standard likelihood are statistically indistinguishable, as shown in Fig. 11. However, with 24 nodes comprising two Intel Xeon E5-2650v4 12x 2.20 GHz 12-Core CPU each, the sampling run times and relative speedup are: $t_{\text{samp}}^{\text{ROQ}=0}/t_{\text{samp}}^{\text{ROQ}=1} = 49h20m/4h14m \sim 11.3$, in very good agreement with the predictions presented above. Presampling interpolant construction took nine mins per detector with these settings.

Finally, we stress that the speedup resulting from the combination of `mlgw_bns` and ROQ will bear a more dramatic impact when applied to longer frequency axes. For example, in the case of full inspiral-merger-postmerger BNS signals analyses, with a lower frequency bound of ~ 5 Hz and reaching up to ~ 8 kHz, applications of similar techniques will provide a speedup larger than three orders of magnitude compared to a uniform grid.

V. CONCLUSIONS

In this work we have introduced `mlgw_bns`, a cutting-edge ML surrogate waveform approximant in the frequency domain for spin-aligned BNS mergers, designed for applications to both current and future GW detectors. Our model is trained on highly accurate `TEOBResumSPA` EOB waveforms, faithfully represented with a fidelity larger than the accuracy of the baseline SPA model against the native time-domain EOB model ($\bar{\mathcal{F}} \lesssim 10^{-5}$). At the same time, thanks to several dimensional reductions steps, `mlgw_bns` delivers a remarkable increase in efficiency, achieving the fastest waveform generation among all the available approximants including comparable physical information.

By performing careful benchmark tests with varying frequency grids, we estimate a speedup of ~ 30 with respect to `TEOBResumSPA`, when evaluated on frequency axes composed of less than $\sim 10^4$ points, which can reach up to ~ 35 for less than $\sim 10^2$ frequency points. Combined with ROQ techniques, an overall PE acceleration of more than an order of magnitude is achieved for current BNS analyses, as we explicitly demonstrated reanalyzing GW170817 using a reduced basis. Thanks to the improved performance of our model, in our investigations the likelihood cost is no longer dominated by the waveform generation time, but by inner products' computations, making additional decreases in the evaluation time of our ML model less relevant. If the inner products' computation cost can be reduced in future PE implementations, it will be important to explore further optimizations of the algorithm, such as tuning the number of PCA components

and the greedy downsampling reconstruction tolerance, or improving the hyperparameters selection procedure.

Analysis targeting next generation detectors' observations, such as PE studies in the ET band, will benefit from even more dramatic improvements. Since the number of empirical nodes will still be $O(10^2)$ even at high SNR [5], well into the plateaux of Fig. 8, ROQ interpolants will allow keeping the waveform generation cost almost identical to the one of current analyses. This in turn will lead to a waveform generation speedup of ~ 50, compared to a standard uniform grid when analyzing a signal starting from 5 Hz. Instead, given the extremely low overhead of our ML model, the combined usage of ROQ and `mlgw_bns` will provide a massive speedup of more than $\sim 10^3$ for the same configuration, without loss of accuracy. Finally, thanks to the usage of PN-hybridization techniques, our model can be evaluated from arbitrarily low frequencies, readily allowing analyses exploring the discovery potential of planned space-based deci-Hertz detectors.

Other than exploiting fast PE techniques, our ML model can even enable them. In fact, posterior sampling acceleration through, e.g., the application of Hamiltonian nested sampling [84], as well as forecasting with Fisher matrix studies, can be easily achieved thanks to the intrinsically differentiable architecture of `mlgw_bns`: a planned neural network upgrade is to yield not only the waveform polarizations $h_{+,\times}$ but also their derivatives with respect to the parameters, i.e., $\partial h_{+,\times}/\partial \theta_i$. The knowledge of gradients can be also exploited in template bank generation [85], allowing for a fast computation of a metric approximation for the match and for coverage of a large dimensional parameter space: our model will facilitate the generation of the first BNS template bank including tidal effects.

In the future, the baseline model and the physics content of `mlgw_bns` will require improvements in order to meet the accuracy prerequisites of XG observatories. While the simplicity in retraining `mlgw_bns` will allow it to remain up to date with future enhancements of tidal EOB models (such as self-spin interactions, higher order tidal effects, dynamical tides), new challenges will be represented by the inclusion of: higher modes [86], precession [87,88], eccentricity [89,90],[7] and a frequency-domain postmerger completion [33,58]. We leave such extensions of `mlgw_bns` to future work, but briefly discuss possible strategies to tackle them. Higher order ($\ell > 2$) modes break the simple dependence on the inclination angle ι described by Eq. (3), requiring the modes to be reconstructed separately, with a corresponding slowdown in waveform evaluation. Graphical processing unit acceleration [45] could be employed to ameliorate this. Precession effects could be immediately included relying on ML-reconstruced higher modes, and

---

[7]All these features are already implemented both in the native time-domain `TEOBResumS` model and in `TEOBResumSPA`, with the exclusion of eccentricity, only available in the time-domain waveform.

subsequently applying a twisting [88,91,92] describing generic spin dynamics. Finally, eccentricity introduces modulations which make the time-to-frequency map non-monotonic; this prevents a straightforward application of SPA, which we use to generate our training datasets. This problem could be cured by moving from SPA to shifted uniform asymptotics [93,94].

In summary, `mlgw_bns` enables an important leap towards feasible and accurate PE with XG detectors, immediately providing a very efficient alternative to current EOB BNS models for present-day analyses.

APPENDIX A: BENCHMARKING PROCEDURE

In this section we detail the procedure used in order to compute the benchmarks shown in Fig. 8, and we draw a comparison to other existing EOB surrogates.
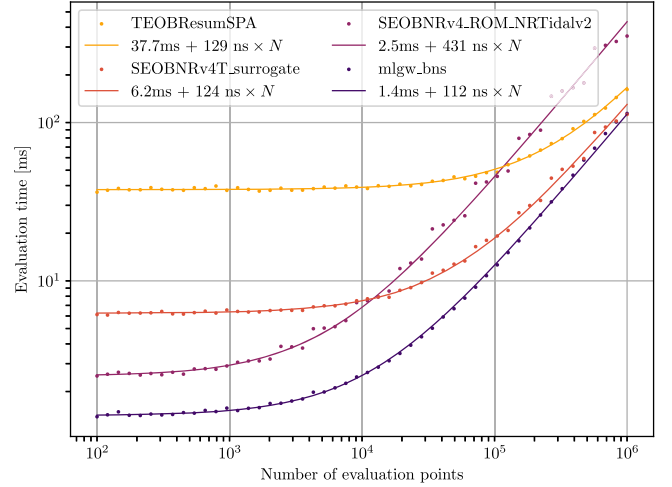

FIG. 12. Evaluation times for waveforms starting at 15 Hz for several approximants. `SEOBNRv4_ROM_NRTidalv2` and `SEOBNRv4T_surrogate` are called through their LAL Python interfaces.

The times we are measuring are quite short and susceptible to fluctuations, therefore we must average over several trials. Further, the order in which tests are performed may affect the results, therefore we randomize the trials across several epochs.

We create a batch of test cases with all possible combinations of the following:
(1) the N_{app} relevant approximants;
(2) the N_{grid} relevant frequency grid sizes n_{freq}, chosen to be integer approximations of a logarithmically spaced grid; and
(3) N_{seed} seeds for the parameters—for each seed a single set of parameters is generated.

For each test case, we define a uniformly spaced[8] frequency grid with n_{freq} points between f_{min} and 2048 Hz. In Fig. 8 we chose $f_{min} = 5$ Hz, while in this section we choose $f_{min} = 15$ Hz in order to compare with other models, and the results can be seen in Fig. 12. We then randomly generate a single set of parameters, with the same procedure used for the generation of the model and with a different seed. For the generation of Figs. 8 and 12 we use $N_{grid} = 50$, $N_{seed} = 20$, and run all tests for ten epochs, shuffling them each time.

The `mlgw_bns` model used in Fig. 12 is not the exact same one whose performance is discussed in the rest of the paper; the higher initial frequency allows us to use a smaller reduced frequency grid. Besides this change, however, the training parameters are the same, and it was also verified to have mismatches $\lesssim 10^{-5}$ with its training reference.

---

[8]As discussed in Sec. II C 1 this is not a good choice for PE, but this is not an issue since for the purposes of benchmarking the number of points in the frequency grid is the only relevant parameter.

APPENDIX B: HYPERPARAMETERS USED

The hyperparameters used for the network whose performance is discussed in this work are as follows:

(1) the exponent α for the principal components is set to 0.37;

(2) the network has two hidden layers with sizes 82 and 95 respectively;

(3) the activation function chosen is tanh;

(4) the L2 regularization parameter is set to 10^{-4};

(5) the batch size for the Adam algorithm is set to 162;

(6) the initial learning rate is set to 1.2×10^{-3};

(7) the fraction of data kept for internal validation during the training is 7.5% (this refers to the internal validation step performed automatically by SCIKIT-LEARN); and

(8) when the accuracy does not improve by more than 2×10^{-15} for 91 epochs the training stops. Note that this number refers to Euclidean distances across PCA-reduced residuals, so it has no direct physical meaning.

The downsampling indices are determined by training on a dataset of $2^{10} = 1024$ waveforms, the PCA is trained on $2^{16} = 65536$ waveforms, the network is trained on $2^{17} = 131072$ waveform residuals. All these training datasets are independent, extracted from the same uniform distribution in the ranges of the parameters.

APPENDIX C: ROQ CONSTRUCTION

1. `PyROQ` summary

We first give a brief recap of the `PyROQ` algorithm presented in Ref. [83] (see their "Algorithm 1"), to allow for an easier comparison with the applied software improvements and differences in our training strategy; we refer the reader there for a more detailed description. In the first step of the `PyROQ` algorithm, a "preselected" dataset of waveform vectors (typically referred to as "basis") is constructed. This preselected basis is initially composed of waveform vectors evaluated at the corner of the parameter space.[9] Corner elements are then augmented by randomly generating a waveform dataset of 10^5 elements, and adding to the basis the element with the largest residuals after projection onto the basis. This is repeated until a user-specified tolerance is reached. In a second step, the preselected basis is enriched by generating four increasingly larger datasets of $10^4, 10^5, 10^6, 5 \times 10^6$ elements, which might have different tolerance thresholds. In each

---

[9]Where the corners are not all the ones of the parameter space N-cube, but simply the two points consisting of the lower and upper bounds of each parameter. Also in our algorithm below, we chose to follow the same strategy, since we found that including all the N-dimensional cube edges of the physical parameter space leads to many repetitions which inefficiently increase the number of total basis elements.

of these datasets, the element with the largest interpolation error is added onto the basis, iterating until all the elements of the dataset can be represented with a given accuracy.

2. `JenpyROQ` summary

In this section we discuss our updated version of `PyROQ`, called `JenpyROQ`. Compared to `PyROQ`, we allow for arbitrary enrichment cycles to be defined by the user, add a few strategies to avoid ill-conditioning of the interpolants construction (which is always monitored as a basic sanity check), and perform a restructuring of the code. The latter consists of: imposing modularity at all stages; implementing MPI parallelization, generalizing the parameter handling (useful when dealing with flexible models with many more parameters than standard binaries, e.g., BNS post-merger models [32,33,58]); extending I/O management; and adding a generic waveform class capable of interfacing with modern Python-based models.

Regarding the algorithm itself (again, using as a reference template "Algorithm 1" of [83]), in the preselection cycle, we complement the initial corner basis until the total basis size reaches either a given number of elements N_{pre} or a specified tolerance ϵ_{pre} (which might be different from the final target tolerance ϵ_{ROQ}). At each preselection step j, with $j = 1, \ldots, N_{\mathrm{pre}} - 2$, we generate N_{step} new random waveforms, among which we pick the element worst represented by projection on the current basis. Later, in the enrichment steps, we adopt a more flexible scheme and perform an arbitrary number of enrichment cycles N_{cycles}, each with arbitrary size L^i, number of outliers N_{out}^i, and threshold ϵ^i, with $i = 1, \ldots, N_{\mathrm{cycles}}$. Finally, we test the constructed interpolants on N_{out} out-of-training datapoints.

By imposing a maximum N_{step} in the preselection phase (which can typically be guessed from rapid exploratory investigations), we keep the size of the preselection datasets ($N_{\mathrm{pre}} \times N_{\mathrm{step}}$) small, and so ensure that only a small number of evaluations are performed until the size of the basis is already large.[10] Only after the preselected basis has a substantial number of elements, we compare against increasingly larger datasets in the enrichment steps. This way, most of the elements are already accurately represented, and no need of re-evaluating many comparisons arises, allowing us to keep a small memory footprint. Furthermore, using a stricter relative tolerance ($\epsilon^i/\epsilon_{\mathrm{ROQ}}$) in the initial enrichment cycles ensures that an accurate interpolant is constructed with only a small number of computations. Given the exponential convergence of the algorithm, such accuracy typically carries over subsequent, much larger datasets at the true target tolerance. The

---

[10]This does not apply to the cases where ϵ_{pre} is reached, but in such a case only a small number of subsequent evaluations on larger datasets are expected if $\epsilon_{\mathrm{pre}} < \epsilon_{\mathrm{ROQ}}$, which we typically impose.

constructed interpolants are interfaced with the Bajes pipeline [38], following the likelihood formulation of [60]. Future planned developments include adaptive frequency sampling, on the lines of [60] and of Sec. II C 1 of this work.

Finally, although the interpolants construction is always formally well defined because the basis matrix in Eq. 16 of [39] is always invertible,[11] in some cases the numerical inversion of the basis matrix can be ill conditioned. This happens when the algorithm finds the same empirical interpolant point more than once; they have to be unique to avoid double counting of frequencies in the likelihood. Ill conditioning is more frequent for longer waveforms (corresponding to low-mass binaries), and it is caused by large disparities in the module of the basis matrix eigenvalues, implying a high conditioning number (when assuming a Euclidean norm), hence larger numerical instabilities. Ill conditioning appeared in early attempts to construct basis relevant to this work as soon as the segment length was increasing above $T = 128$ s. An easy solution we have found to this problem is simply to avoid applying a Gram-Schmidt procedure to the new basis elements (which are only strictly required to be independent, not orthonormal [39]), using waveform vectors as basis elements as opposed to Gram-Schmidt residuals [83]. This avoids sparsity in the matrix construction, significantly decreasing the conditioning number. For our current purposes, this simple fix was sufficient to ensure numerical stability. In case ill conditioning is still present (a situation which we have not encountered yet in our explorations), currently the algorithm is forced to proceed, simply discarding the repeated interpolation point and switching to the interpolation point with the second worst residual. We expect this latter procedure to slightly decrease the efficiency of the basis construction (i.e., to increase the number of elements required to reach the required precision), but given the exponential convergence of the

_____

[11]Except for the trivial distance rescaling that may be factored out in ROQ constructions, waveforms of binaries in GR are always linearly independent. This discussion does not apply to models which include exact degeneracies among basis elements when varying the physical parameters.

algorithm, a small increase in the number of required elements does not appreciably affect PE run times. In the future, we plan to implement a more robust solution to the above issues by exploring modifications of our naive Gram-Schmidt algorithm, along the lines of what was discussed in Appendix A of [39] and their Refs. [66–68]. In any case, the matrix basis conditioning number is always monitored in our algorithm, and used to flag an ill-posed basis construction.

3. Details of GW170817 interpolants construction

We train the ROQ basis on the full range of validity of `mlgw_bns`, except for the chirp mass, which is only varied within the range $\mathcal{M}_c \in [1.1968, 1.1988]$, wide enough to safely encompass the full posterior of GW170817. This is not restrictive; the chirp mass can already be reliably estimated within an even tighter accuracy interval by low-latency GW searches of BNS signals [100], hence a similar "targeted" basis could be constructed even for realistic, new data. We set a tolerance threshold of $10^{-4}(10^{-6})$ for the linear (quadratic) basis, ensuring that the ROQ interpolants are valid up to an SNR of $O(70\text{–}100)$, see, e.g., Eq. 8 of Ref. [101]. We use a total of 4.1×10^5 training data points, split between the preselection and the enrichment steps. A preselected basis is constructed using $N_{\text{pre}} = 200(10)$ elements for the linear (quadratic) case, and $N_{\text{step}} = 1000$ points at each step. We set three enrichment cycles, each composed of $[10^4, 10^5, 10^5]$ data points, $N_{\text{out}}^i = 0$, and a respective relative tolerance of $[0.1, 1.0, 1.0]$. The resulting bases are composed of 267 (10) linear (quadratic) elements, achieving a linear (quadratic) frequency axis reduction factor of 950 (25300). We confirm the accuracy of the constructed interpolants by applying them to the reconstruction of 5×10^5 validation data points, finding less than 0.1% outliers and none above the 10^{-3} threshold. Direct evaluation of 5×10^5 likelihood inner products confirmed the above tests, always showing errors smaller than the requested accuracy threshold.

The configuration file used to produce ROQ basis employed in this work is available at: https://github.com/GCArullo/JenpyROQ/blob/main/config_files/config_MLGW-BNS_LVK_GW170817_release.ini.

[1] J. Aasi *et al.* (LIGO and Virgo Collaborations), Parameter estimation for compact binary coalescence signals with the first generation gravitational-wave detector network, Phys. Rev. D **88**, 062001 (2013).

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

[3] A. Buonanno, B. Iyer, E. Ochsner, Y. Pan, and B. Sathyaprakash, Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors, Phys. Rev. D **80**, 084043 (2009).

[4] L. Blanchet, Gravitational radiation from post-Newtonian sources and inspiralling compact binaries, Living Rev. Relativity **17,** 2 (2014).

[5] R. Smith *et al.*, Bayesian Inference for Gravitational Waves from Binary Neutron Star Mergers in Third Generation Observatories, Phys. Rev. Lett. **127,** 081102 (2021).

[6] G. Pratten, P. Schmidt, and T. Hinderer, Gravitational-wave asteroseismology with fundamental modes from compact binary inspirals, Nat. Commun. **11,** 2553 (2020).

[7] N. Williams, G. Pratten, and P. Schmidt, Prospects for distinguishing dynamical tides in inspiralling binary neutron stars with third generation gravitational-wave detectors, Phys. Rev. D **105,** 123032 (2022).

[8] 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).

[9] P. Schmidt and T. Hinderer, Frequency domain model of f-mode dynamic tides in gravitational waveforms from compact binary inspirals, Phys. Rev. D **100,** 021501 (2019).

[10] R. Gamba, M. Breschi, S. Bernuzzi, M. Agathos, and A. Nagar, Waveform systematics in the gravitational-wave inference of tidal parameters and equation of state from binary neutron star signals, Phys. Rev. D **103,** 124015 (2021).

[11] A. Buonanno and T. Damour, Effective one-body approach to general relativistic two-body dynamics, Phys. Rev. D **59,** 084006 (1999).

[12] A. Buonanno and T. Damour, Transition from inspiral to plunge in binary black hole coalescences, Phys. Rev. D **62,** 064015 (2000).

[13] T. Damour, P. Jaranowski, and G. Schaefer, On the determination of the last stable orbit for circular general relativistic binaries at the third post-Newtonian approximation, Phys. Rev. D **62,** 084011 (2000).

[14] T. Damour, Coalescence of two spinning black holes: An effective one- body approach, Phys. Rev. D **64,** 124013 (2001).

[15] T. Damour, P. Jaranowski, and G. Schäfer, Effective one body approach to the dynamics of two spinning black holes with next-to-leading order spin-orbit coupling, Phys. Rev. D **78,** 024009 (2008).

[16] T. Damour and A. Nagar, An improved analytical description of inspiralling and coalescing black-hole binaries, Phys. Rev. D **79,** 081503 (2009).

[17] T. Damour and A. Nagar, New effective-one-body description of coalescing nonprecessing spinning black-hole binaries, Phys. Rev. D **90,** 044018 (2014).

[18] A. Bohé, L. Shao, A. Taracchini, A. Buonanno, S. Babak, I. Harry, I. Hinder, S. Ossokine, M. Purrer, V. Raymond, T. Chu, H. Fong, P. Kumar, H. Pfeiffer, M. Boyle, D. Hemberger, L. Kidder, G. Lovelace, M. Scheel, and B. Szilagyi, Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors, Phys. Rev. D **95,** 044028 (2017).

[19] A. Nagar *et al.*, Time-domain effective-one-body gravitational waveforms for coalescing compact binaries with nonprecessing spins, tides and self-spin effects, Phys. Rev. D **98,** 104052 (2018).

[20] B. D. Lackey, M. Prrer, A. Taracchini, and S. Marsat, Surrogate model for an aligned-spin effective one body waveform model of binary neutron star inspirals using Gaussian process regression, Phys. Rev. D **100,** 024002 (2019).

[21] T. Damour, B. R. Iyer, and A. Nagar, Improved resummation of post-Newtonian multipolar waveforms from circularized compact binaries, Phys. Rev. D **79,** 064004 (2009).

[22] G. Riemenschneider, P. Rettegno, M. Breschi, A. Albertini, R. Gamba, S. Bernuzzi, and A. Nagar, Assessment of consistent next-to-quasicircular corrections and postadiabatic approximation in effective-one-body multipolar waveforms for binary black hole coalescences, Phys. Rev. D **104,** 104045 (2021).

[23] A. Albertini, A. Nagar, P. Rettegno, S. Albanesi, and R. Gamba, Waveforms and fluxes: Towards a self-consistent effective one body waveform model for nonprecessing, coalescing black-hole binaries for third generation detectors, Phys. Rev. D **105,** 084025 (2022).

[24] E. E. Flanagan and T. Hinderer, Constraining neutron star tidal Love numbers with gravitational wave detectors, Phys. Rev. D **77,** 021502 (2008).

[25] T. Damour and A. Nagar, Effective one body description of tidal effects in inspiralling compact binaries, Phys. Rev. D **81,** 084016 (2010).

[26] D. Bini, T. Damour, and G. Faye, Effective action approach to higher-order relativistic tidal interactions in binary systems and their effective one body description, Phys. Rev. D **85,** 124034 (2012).

[27] S. Bernuzzi, A. Nagar, M. Thierfelder, and B. Brügmann, Tidal effects in binary neutron star coalescence, Phys. Rev. D **86,** 044030 (2012).

[28] S. Bernuzzi, T. Dietrich, and A. Nagar, Modeling the Complete Gravitational Wave Spectrum of Neutron Star Mergers, Phys. Rev. Lett. **115,** 091101 (2015).

[29] T. Hinderer *et al.*, Effects of Neutron-Star Dynamic Tides on Gravitational Waveforms Within the Effective-One-Body Approach, Phys. Rev. Lett. **116,** 181101 (2016).

[30] J. Steinhoff, T. Hinderer, A. Buonanno, and A. Taracchini, Dynamical tides in general relativity: Effective action and effective-one-body Hamiltonian, Phys. Rev. D **94,** 104028 (2016).

[31] S. Akcay, S. Bernuzzi, F. Messina, A. Nagar, N. Ortiz, and P. Rettegno, Effective-one-body multipolar waveform for tidally interacting binary neutron stars up to merger, Phys. Rev. D **99,** 044051 (2019).

[32] M. Breschi, S. Bernuzzi, F. Zappa, M. Agathos, A. Perego, D. Radice, and A. Nagar, KiloHertz gravitational waves from binary neutron star remnants: Time-domain model and constraints on extreme matter, Phys. Rev. D **100,** 104029 (2019).

[33] M. Breschi, S. Bernuzzi, K. Chakravarti, A. Camilletti, A. Prakash, and A. Perego, Kilohertz gravitational waves from binary neutron star mergers: Numerical-relativity informed postmerger model, arXiv:2205.09112.

[34] A. Nagar and P. Rettegno, Efficient effective one body time-domain gravitational waveforms, Phys. Rev. D **99,** 021501 (2019).

[35] J. Aasi *et al.* (LIGO Scientific Collaboration), Advanced LIGO, Classical Quantum Gravity **32,** 074001 (2015).

[36] F. Acernese *et al.* (VIRGO Collaboration), Advanced Virgo: A second-generation interferometric gravitational wave detector, Classical Quantum Gravity **32,** 024001 (2015).

[37] R. Gamba, S. Bernuzzi, and A. Nagar, Fast, faithful, frequency-domain effective-one-body waveforms for compact binary coalescences, Phys. Rev. D **104,** 084058 (2021).

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

[39] 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).

[40] M. Tiglio and A. Villanueva, Reduced order and surrogate models for gravitational waves, Living Rev. Relativity **25,** 2 (2022).

[41] A. J. K. Chua, C. R. Galley, and M. Vallisneri, Reduced-Order Modeling with Artificial Neurons for Gravitational-Wave Inference, Phys. Rev. Lett. **122,** 211101 (2019).

[42] S. Khan and R. Green, Gravitational-wave surrogate models powered by artificial neural networks, Phys. Rev. D **103,** 064015 (2021).

[43] S. Schmidt, M. Breschi, R. Gamba, G. Pagano, P. Rettegno, G. Riemenschneider, S. Bernuzzi, A. Nagar, and W. Del Pozzo, Machine learning gravitational waves from binary black hole mergers, Phys. Rev. D **103,** 043020 (2021).

[44] D. Barsotti, F. Cerino, M. Tiglio, and A. Villanueva, Gravitational wave surrogates through automated machine learning, Classical Quantum Gravity **39,** 085011 (2022).

[45] L. M. Thomas, G. Pratten, and P. Schmidt, Accelerating multimodal gravitational waveforms from precessing compact binaries with artificial neural networks, Phys. Rev. D **106,** 104029 (2022).

[46] C.-H. Liao and F.-L. Lin, Deep generative models of gravitational waveforms via conditional autoencoder, Phys. Rev. D **103,** 124051 (2021).

[47] B. D. Lackey, S. Bernuzzi, C. R. Galley, J. Meidam, and C. Van Den Broeck, Effective-one-body waveforms for binary neutron stars using surrogate models, Phys. Rev. D **95,** 104036 (2017).

[48] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral, Phys. Rev. Lett. **119,** 161101 (2017).

[49] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs, Phys. Rev. X **9,** 031040 (2019).

[50] M. Favata, Systematic Parameter Errors in Inspiraling Neutron Star Binaries, Phys. Rev. Lett. **112,** 101101 (2014).

[51] T. Damour, A. Nagar, and L. Villain, Measurability of the tidal polarizability of neutron stars in late-inspiral gravitational-wave signals, Phys. Rev. D **85,** 123007 (2012).

[52] P. Ajith, M. Boyle, D. A. Brown, S. Fairhurst, M. Hannam, I. Hinder, S. Husa, B. Krishnan, R. A. Mercer, F. Ohme, C. D. Ott, J. S. Read, L. Santamaria, and J. T. Whelan, Data formats for numerical relativity waves, arXiv:0709.0093.

[53] E. S. Team, Einstein gravitational wave telescope conceptual design study, ET-0106C-10, 10.5281/zenodo.3911261 (2011).

[54] S. Hild *et al.*, Sensitivity studies for third-generation gravitational wave observatories, Classical Quantum Gravity **28,** 094013 (2011).

[55] F. Messina, R. Dudi, A. Nagar, and S. Bernuzzi, Quasi-5.5PN `TaylorF2` approximant for compact binaries: Point-mass phasing and impact on the tidal polarizability inference, Phys. Rev. D **99,** 124051 (2019).

[56] Q. Henry, G. Faye, and L. Blanchet, Tidal effects in the gravitational-wave phase evolution of compact binary systems to next-to-next-to-leading post-Newtonian order, Phys. Rev. D **102,** 044033 (2020).

[57] A. Nagar, F. Messina, P. Rettegno, D. Bini, T. Damour, A. Geralico, S. Akcay, and S. Bernuzzi, Nonlinear-in-spin effects in effective-one-body waveform models of spin-aligned, inspiralling, neutron star binaries, Phys. Rev. D **99,** 044007 (2019).

[58] M. Breschi, R. Gamba, S. Borhanian, G. Carullo, and S. Bernuzzi, Kilohertz gravitational waves from binary neutron star mergers: Inference of postmerger signals with the Einstein Telescope, arXiv:2205.09979.

[59] S. Vinciguerra, J. Veitch, and I. Mandel, Accelerating gravitational wave parameter estimation with multi-band template interpolation, Classical Quantum Gravity **34,** 115006 (2017).

[60] R. Smith, S. E. Field, K. Blackburn, C.-J. Haster, M. Prrer, V. Raymond, and P. Schmidt, Fast and accurate inference on gravitational waves from precessing compact binaries, Phys. Rev. D **94,** 044031 (2016).

[61] M. Maggiore, *Gravitational Waves. Vol. 1: Theory and Experiments*, Oxford Master Series in Physics (Oxford University Press, New York, 2007).

[62] J. A. Clark, A. Bauswein, N. Stergioulas, and D. Shoemaker, Observing gravitational waves from the post-merger phase of binary neutron star coalescence, Classical Quantum Gravity **33,** 085003 (2016).

[63] P. J. Easter, S. Ghonge, P. D. Lasky, A. R. Casey, J. A. Clark, F. H. Vivanco, and K. Chatziioannou, Detection and parameter estimation of binary neutron star merger remnants, Phys. Rev. D **102,** 043011 (2020).

[64] T. Soultanis, A. Bauswein, and N. Stergioulas, Analytic models of the spectral properties of gravitational waves from neutron star merger remnants, Phys. Rev. D **105,** 043020 (2022).

[65] M. Wijngaarden, K. Chatziioannou, A. Bauswein, J. A. Clark, and N. J. Cornish, Probing neutron stars with the full pre-merger and post-merger gravitational wave signal from binary coalescences, Phys. Rev. D **105,** 104019 (2022).

[66] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Édouard

Duchesnay, SCIKIT-LEARN: Machine learning in PYTHON, J. Mach. Learn. Res. **12,** 2825 (2011).

[67] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.

[68] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, OPTUNA: A next-generation hyperparameter optimization framework, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery; Data Mining* (Association for Computing Machinery, New York, NY, USA, 2019), pp. 2623–2631, 10.1145/3292500.3330701.

[69] Y. Ozaki, Y. Tanigaki, S. Watanabe, and M. Onishi, Multiobjective tree-structured parzen estimator for computationally expensive optimization problems, in *Proceedings of the 2020 Genetic and Evolutionary Computation Conference, GECCO '20* (Association for Computing Machinery, New York, NY, USA, 2020), pp. 533–541, 10.1145/3377930.

[70] S. K. Lam, A. Pitrou, and S. Seibert, NUMBA: A LLVM-based PYTHON JIT compiler, in *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC, LLVM '15* (Association for Computing Machinery, 2015), pp. 1–6, 10.1145/2833157.2833162.

[71] B. Zackay, L. Dai, and T. Venumadhav, Relative binning and fast likelihood evaluation for gravitational wave parameter estimation, arXiv:1806.08792.

[72] 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).

[73] J. S. Speagle, DYNESTY: A dynamic nested sampling package for estimating Bayesian posteriors and evidences, Mon. Not. R. Astron. Soc. **493,** 3132 (2020).

[74] J. Veitch and A. Vecchio, Bayesian coherent analysis of inspiral gravitational wave signals with a detector network, Phys. Rev. D **81,** 062003 (2010).

[75] E. Thrane and C. Talbot, An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models, Publ. Astron. Soc. Aust. **36,** e010 (2019).

[76] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Properties of the Binary Neutron Star Merger GW170817, Phys. Rev. X **9,** 011001 (2019).

[77] S. E. Field, C. R. Galley, F. Herrmann, J. S. Hesthaven, E. Ochsner, and M. Tiglio, Reduced Basis Catalogs for Gravitational Wave Templates, Phys. Rev. Lett. **106,** 221102 (2011).

[78] S. Caudill, S. E. Field, C. R. Galley, F. Herrmann, and M. Tiglio, Reduced basis representations of multi-mode black hole ringdown gravitational waves, Classical Quantum Gravity **29,** 095016 (2012).

[79] H. Antil, S. E. Field, F. Herrmann, R. H. Nochetto, and M. Tiglio, Two-step greedy algorithm for reduced order quadratures, J. Sci. Comput. **57,** 604 (2013).

[80] 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).

[81] P. Canizares, S. E. Field, J. Gair, V. Raymond, R. Smith, and M. Tiglio, Accelerated Gravitational-Wave Parameter Estimation with Reduced Order Modeling, Phys. Rev. Lett. **114,** 071104 (2015).

[82] J. Meidam, K. Tsang, J. Goldstein, M. Agathos, A. Ghosh, C. Haster, V. Raymond, A. Samajdar, P. Schmidt, R. Smith, K. Blackburn, W. DelPozzo, S. Field, T. Li, M. Purrer, C. VanDenBroeck, J. Veitch, and S. Vitale, Parametrized tests of the strong-field dynamics of general relativity using gravitational wave signals from coalescing binary black holes: Fast likelihood calculations and sensitivity of the method, Phys. Rev. D **97,** 044033 (2018).

[83] H. Qi and V. Raymond, PYTHON-based reduced order quadrature building code for fast gravitational wave inference, Phys. Rev. D **104,** 063031 (2021).

[84] M. Betancourt, Nested sampling with constrained Hamiltonian Monte Carlo, AIP Conf. Proc. **1305,** 165 (2011).

[85] A. Coogan, T. D. P. Edwards, H. S. Chia, R. N. George, K. Freese, C. Messick, C. N. Setzer, C. Weniger, and A. Zimmerman, Efficient template bank generation with differentiable waveforms, Phys. Rev. D **106,** 122001 (2022).

[86] A. Nagar, G. Riemenschneider, G. Pratten, P. Rettegno, and F. Messina, Multipolar effective one body waveform model for spin-aligned black hole binaries, Phys. Rev. D **102,** 024077 (2020).

[87] S. Akcay, R. Gamba, and S. Bernuzzi, A hybrid post-Newtonian—effective-one-body scheme for spin-precessing compact-binary waveforms, Phys. Rev. D **103,** 024014 (2021).

[88] R. Gamba, S. Akçay, S. Bernuzzi, and J. Williams, Effective-one-body waveforms for precessing coalescing compact binaries with post-Newtonian twist, Phys. Rev. D **106,** 024020 (2022).

[89] D. Chiaramello and A. Nagar, Faithful analytical effective-one-body waveform model for spin-aligned, moderately eccentric, coalescing black hole binaries, Phys. Rev. D **101,** 101501 (2020).

[90] A. Nagar, A. Bonino, and P. Rettegno, Effective one-body multipolar waveform model for spin-aligned, quasicircular, eccentric, hyperbolic black hole binaries, Phys. Rev. D **103,** 104021 (2021).

[91] P. Schmidt, M. Hannam, S. Husa, and P. Ajith, Tracking the precession of compact binaries from their gravitational-wave signal, Phys. Rev. D **84,** 024046 (2011).

[92] P. Schmidt, M. Hannam, and S. Husa, Towards models of gravitational waveforms from generic binaries: A simple approximate mapping between precessing and non-precessing inspiral signals, Phys. Rev. D **86,** 104063 (2012).

[93] A. Klein, Y. Boetzel, A. Gopakumar, P. Jetzer, and L. de Vittori, Fourier domain gravitational waveforms for precessing eccentric binaries, Phys. Rev. D **98,** 104043 (2018).

[94] A. Klein, N. Cornish, and N. Yunes, Fast frequency-domain waveforms for spin-precessing binary inspirals, Phys. Rev. D **90**, 124029 (2014).

[95] https://www.gw-openscience.org.

[96] github.com/jacopok/mlgw_bns.

[97] github.com/GCArullo/JenpyROQ.

[98] bitbucket.org/eob_ihes/teobresums/.

[99] github.com/matteobreschi/bajes.

[100] S. Biscoveanu, S. Vitale, and C.-J. Haster, The reliability of the low-latency estimation of binary neutron star chirp mass, Astrophys. J. Lett. **884**, L32 (2019).

[101] M. Pürrer and C.-J. Haster, Gravitational waveform accuracy requirements for future ground-based detectors, Phys. Rev. Res. **2**, 023151 (2020).