

Numerical relativity surrogate waveform models for exotic compact objects: The case of head-on mergers of equal-mass Proca stars

Raimon Luna^{1,*}, Miquel Llorens-Monteaudo¹, Ana Lorenzo-Medina^{2,†}, Juan Calderón Bustillo^{2,3,‡},
Nicolas Sanchis-Gual¹, Alejandro Torres-Forné^{1,4}, José A. Font^{1,4},
Carlos A. R. Herdeiro⁵ and Eugen Radu⁵

¹*Departamento de Astronomía y Astrofísica, Universitat de València,
Dr. Moliner 50, 46100, Burjassot (València), Spain*

²*Instituto Galego de Física de Altas Enerxías, Universidade de Santiago de Compostela,
15782 Santiago de Compostela, Galicia, Spain*

³*Department of Physics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong*

⁴*Observatori Astronòmic, Universitat de València, C/ Catedrático José Beltrán 2,
46980 Paterna (València), Spain*

⁵*Departamento de Matemática da Universidade de Aveiro and Centre for Research
and Development in Mathematics and Applications (CIDMA),
Campus de Santiago, 3810-193 Aveiro, Portugal*



(Received 6 April 2024; accepted 5 June 2024; published 16 July 2024)

We present several high-accuracy surrogate models for gravitational-wave signals from equal-mass head-on mergers of Proca stars, computed through the Newman-Penrose scalar ψ_4 . We also discuss the current state of the model extensions to mergers of Proca stars with different masses, and the particular challenges that these present. The models are divided in two main categories: two-stage and monolithic. In the two-stage models, a dimensional reduction algorithm is applied to embed the data in a reduced feature space, which is then interpolated in terms of the physical parameters. For the monolithic models, a single neural network is trained to predict the waveform from the input physical parameter. Our model displays mismatches below 10^{-3} with respect to the original numerical waveforms. Finally, we demonstrate the usage of our model in full Bayesian parameter inference through the accurate recovery of numerical relativity signals injected in zero noise, together with the analysis of GW190521. For the latter, we observe excellent agreement with existing results that make use of full numerical relativity.

DOI: [10.1103/PhysRevD.110.024004](https://doi.org/10.1103/PhysRevD.110.024004)

I. INTRODUCTION

The detection of gravitational waves by the LIGO-Virgo-KAGRA (LVK) detector network has opened a new era of gravitational wave astrophysics [1–9]. During the first three observing runs these detectors have already detected a large collection of signals, $\mathcal{O}(100)$, from binary mergers of compact objects, namely black holes and neutron stars, together with an estimation of their physical parameters. These signals have been mainly detected and studied through the matched-filtering of the experimental data [10–18] with theoretical waveform template banks [12,19,20].

Even though templates from the coalescence of black holes and neutron stars have dominated the search landscape as candidates for gravitational wave sources, the possibility of alternative compact objects, known generically as exotic compact objects (ECO), has generated an

increasing interest in recent years [21]. In particular, much scientific attention has been oriented towards those objects that could generate gravitational-wave signals resembling those of black holes, which could generate confusion in their classification. These objects are typically known as “black hole mimickers.”

A particularly important class of ECOs are those based on gravitationally bound states of ultralight bosonic fields [22,23]. These fields could appear in particle physics models, such as the string axiverse [24,25] or as extensions of the Standard Model [26]. Ultralight bosonic fields are capable of forming stationary states, resembling hydrogen orbitals [27], which can also be rotating [28,29]. In the case of massive complex vector fields—Proca fields—the objects are known as Proca stars [30].

While sharing some similarities, scalar and Proca bosonic stars have distinct behaviors regarding their formation and stability [31]. In particular, some configurations of rotating bosonic stars exhibit a bar-mode instability [32], which can be prevented through the addition of nonlinear interaction terms in their action [33,34] or considering

*raimonluna@gmail.com

†analorenzo.medina@usc.es

‡juan.calderon.bustillo@gmail.com

multistate bosonic configurations [35]. On the other hand, Proca stars are known to have sufficiently generic mechanisms of formation through the so-called gravitational cooling mechanism [36,37] and a stable ground state [38]. A general review on boson stars can be found in [39].

Numerical evolutions of Proca stars have been performed at the fully nonlinear regime in $3 + 1$ numerical relativity, both with single stars [40] and head-on collisions [41,42]. This has allowed the use of the gravitational waveforms extracted from the simulations to compare with actual detections and perform parameter estimations (PE) [43,44]. In particular, statistical inference on the event GW190521 resulted in a very good agreement, even with a slightly higher Bayes factor than with black hole templates [45].

The standard procedure for Bayesian PE involves sampling over a large number of waveform templates generated with different physical parameters, typically on the order of millions. As numerical relativity simulations are computationally expensive, on the order of thousands of CPU hours per simulation, generating all the templates by direct simulation is not possible at the practical level. For the specific case of binary black hole (BBH) mergers, the development of waveform templates (or “approximants”) based upon approximations to the equations of motion of the two-body problem in general relativity, has already a long history spanning over two decades [46]. For the three main model families covering the whole inspiral-merger-ringdown process, namely SEOBNR [47–50], TEOBRESUMS [51–53], and IMRPHENOM [54–57], waveform templates are either calibrated to numerical relativity using analytical or semi-analytical expressions or they are obtained through the “hybridization” of waveforms from numerical relativity simulations and post-Newtonian approximations.

A fourth and more recent line of work is the construction of surrogate waveform models, first proposed in the seminal work of [58]. Such models have been shown to reach an accuracy comparable to numerical relativity and they will be the focus of our work (albeit in the context of Proca-star mergers). For nonspinning quasicircular BBH mergers up to a mass ratio $q = 8$, the first attempt to build a surrogate model employing waveforms from numerical relativity was performed by [59]. This work was soon followed by the first surrogate model for precessing systems [60], NRSUR4DS2, built from a set of 276 precessing BBH simulations, and by the corresponding extensions to account for unequal masses and generic spins NRSUR7DQ2 [61] and NRSUR7DQ4 [62]. The latter was calibrated with 1528 numerical relativity precessing simulations with mass-ratio $q \leq 4$ and spin magnitudes up to 0.8, covering all harmonics with $l \leq 4$, retaining all intrinsic degrees of freedom of quasicircular BBH systems. NRSUR7DQ4 is the most accurate waveform model for quasicircular BBH with generic spins. This model has been employed to analyze several exceptional events from the LVK collaboration [4,63], to extract measurements of novel observables like

orbital precession [64], gravitational-wave recoils [65,66] or the Chern-Pontryagin pseudoscalar [67].

In addition, Islam *et al.* [68] recently used NRSUR7DQ4 to reanalyze 42 events detected by the LVK collaboration. Moreover, for aligned-spin binaries, a surrogate model for hybridized numerical relativity waveforms was constructed in [69] using 104 simulations, employing effective one body results for hybridizing the orbital frequency and post-Newtonian results for hybridizing the harmonic amplitudes. This was extended in [70] who developed a nonprecessing two-dimensional surrogate model of hybridized waveforms up to mass-ratio $q = 15$, NRHYBSUR2DQ15, in order to build a template waveform for signal GW190814 [71]. It is also worth mentioning that efforts based on machine learning, namely Gaussian process regression, have been used to build waveform surrogate models for both nonprecessing and precessing BBH systems at points of the parameter space not covered by numerical relativity simulations [72,73].

The progress in surrogate waveform models for BBH mergers motivates the development of such models for ECOs, in particular if a systematic search for Proca stars is to be pursued. These models should be able to produce accurate waveform estimates, taking as input the physical parameters of the system, using modest computational resources. Some attempts have been made in this direction by application of generative adversarial networks [74], already producing promising results. In this paper, we perform a first approach to the use of surrogate models for PE in Proca star binaries by restricting to head-on collisions of Proca stars with equal masses. The choice of head-on collisions instead of quasicircular inspirals is entirely motivated by the much higher technical difficulty and computational cost of the Proca inspirals, which makes it challenging to perform numerical-relativity simulations of such systems at the present time.

Therefore, in our case, the system depends on a single physical parameter, the normalized frequency of the system ω/μ_V , where μ_V denotes the mass of the ultralight vector boson forming the stars. For the PE procedure to confidently resolve the physical parameters of the system, the error introduced by the surrogate model must be much lower than the typical differences between templates in the sampling range. For this reason, we analyze a variety of possible surrogate architectures that can be used to fit the simulation data. We compare their performances on a testing set and we finally select the most reliable one for the actual Bayesian analysis. The models in this study are data driven, based entirely on a catalog of simulations [42], with no built-in physical information.

In the first class of models, which we call two-phase architectures, we encode the waveform data into a feature space of lower dimension. This drastically reduces the amount of numerical parameters that need to be fitted, thus reducing the model size. The parameters in the feature space (sometimes called latent space) can then be interpolated in

terms of ω/μ_V . As we have only one independent variable, we use standard cubic spline interpolation. The algorithms that we analyze for the dimensional reduction phase are singular value decomposition, empirical interpolant representation [75] and a deep convolutional autoencoder. The first two algorithms decompose the data as a linear combination of a suitable reduced basis, while the last one uses a nonlinear mapping based on neural networks. The second type of model is a more straightforward approach, which we call monolithic, as a single neural network is used to map the frequency values ω/μ_V to the temporal values of the waveform.

We use our model to perform Bayesian parameter inference on both simulated and real data. In the first case, we perform parameter inference on numerical relativity waveforms injected in zero noise, demonstrating that our surrogate can accurately estimate the parameters of such injections. In the second case, we perform parameter inference on the data containing the signal GW190521, showing that our surrogates faithfully match those obtained through numerical relativity waveforms, presented in [43,45], while using significantly less computational resources. This indicates that this type of data-driven surrogate models are a very promising tool to encapsulate the catalogs of numerical boson-star waveforms in order to be efficiently used in detection data analysis.

The paper is structured as follows. In Sec. II we present the generation of the training data and its preparation before it can be fed to the surrogate models. We also discuss the criteria used for the evaluation and comparison of the different models. Section III describes the different architectures of two-stage models, i.e., those who can be divided into encoder and translator modules. A different type of models, the monolithic surrogates, are exposed in Sec. IV. The results of model evaluations, as well as the Bayesian parameter estimations performed with the models, are discussed in Sec. V. We conclude in Sec. VI.

II. THE DATASET

A. The Proca merger catalog

The models are trained on a catalog of 59 equal-mass simulations from [42], of Proca stars in Einstein-Proca theory

$$\mathcal{S} = \int d^4x \sqrt{-g} \left(\frac{R}{16\pi} - \frac{1}{4} F_{\mu\nu} \bar{F}^{\mu\nu} - \frac{1}{2} \mu^2 A_\mu \bar{A}^\mu \right), \quad (1)$$

where A_μ and $F_{\mu\nu}$ are the Proca potential and field strength, respectively, μ is the Proca field mass parameter, and the bar denotes complex conjugation. We also take $c = G = 1$. The interested reader is addressed to [42] and references therein for details on the formalism. The simulations were performed in the EINSTEIN TOOLKIT [76], which uses the CACTUS framework [77], under the

Baumgarte-Shapiro-Shibata-Nakamura formulation using the MCLACHLAN [78,79] and LEAN [80] thorns. Details on the simulations and construction of initial data can be found in [30,40,41]. The thorn containing the Proca equations was implemented in [81] and is available in [82].

Each simulation produces gravitational waveforms for the Newman-Penrose scalar ψ_4 , which are decomposed in spherical harmonic modes. Our study will be centered on the fundamental $(l, m) = (2, 2)$ and $(l, m) = (2, 0)$ modes. An example of the waveforms, together with their prediction by one of the surrogates presented below, is shown in Fig. 1. Even though the Arnowitt–Deser–Misner (ADM) mass of the initial data may vary due to the construction of the Proca star binaries, the numerical relativity waveforms are rescaled so that $M_{\text{ADM}} = 1$ before they are treated in the surrogate pipelines.

B. Preprocessing of the data

The complex ψ_4 waveforms of the original dataset are subjected to a preprocessing pipeline before they are submitted to the actual model. There are a number of reasons for this. First, the algorithms will always tend to be more effective when the samples are similar to each other. For instance, having large time shifts between otherwise very similar waveforms can unnecessarily increase the cost of the fitting. Second, many machine learning algorithms work best with data of numerical value of order one. As the values of ψ_4 vary usually on the order of 10^{-4} , it is highly advisable to rescale the data before making any learning attempt on it. Additionally, numerical relativity simulations

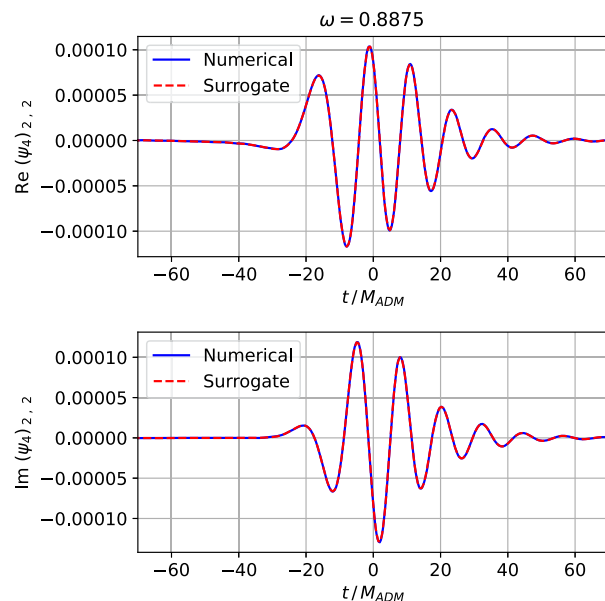


FIG. 1. Example of the reconstruction of a ψ_4 waveform by the singular value decomposition (SVD) surrogate model, for the real (top) and imaginary (bottom) parts of the mode $(l, m) = (2, 2)$ at $\omega = 0.8875$.

usually include a burst of unphysical junk radiation at the beginning of the waveform, which has to be removed. And last but not least, it is highly convenient to have all waveforms on the same time grid, which makes it possible to treat them on equal footing. To this end, a Gaussian curve is fitted to the modulus $|\psi_4|$ of the waves as

$$f(t) = Ae^{-(t-t_0)^2/s^2}. \quad (2)$$

Then, the waveforms are time shifted by t_0 so that all the waves are centered at $t = 0$. Then, they are divided by A so their range of oscillation becomes of order one. In order to perform the transformations, the original waveform is interpolated by cubic splines and reevaluated on a fixed grid of N_ψ equally spaced points between -150 and 150 . ψ_4 is set to zero for $t < 100$, to remove the junk radiation. Figure 2 shows an example of the Gaussian curve fit to the amplitude of a wave after the transformations.

In order for the surrogate to be used for parameter estimation, the relative time positions of the different modes have to be preserved. The values of the time shift and amplitude (t_0, A) are stored in the dataset, as they also need to be interpolated in order to recover the information later. Once the model is evaluated, we multiply the waveforms by their respective amplitudes A and we time shift them so that the peak of the reference mode $(l, m) = (2, 2)$ lies at $t = 0$. This way, the original relative time positions of all the modes are preserved.

The method of the Gaussian fit for the identification of the peak and amplitude of the waveform, as opposed to just taking the absolute maximum of $|\psi_4|$, guarantees that the parameters A and t_0 will vary continuously with ω/μ_V . The absolute maximum of $|\psi_4|$, on the other hand, can suffer discontinuous jumps between relative maxima, which would create discontinuous jumps on the data, thus damaging the performance of the fitting algorithms.

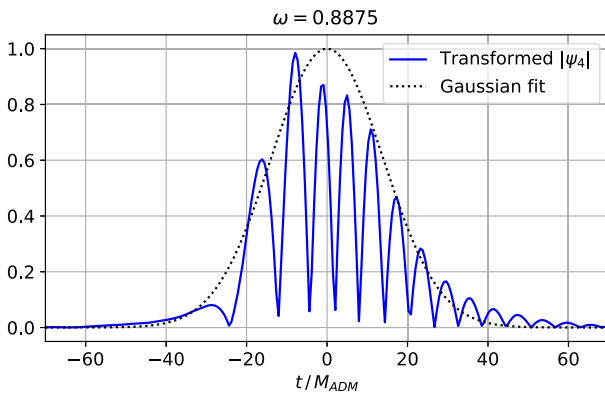


FIG. 2. Transformed profile for the amplitude $|\psi_4|$ of the mode $(l, m) = (2, 2)$ at $\omega = 0.8875$, together with the fitted Gaussian curve. After the transformation, the peak of the Gaussian lies at coordinates $(x, y) = (0, 1)$.

C. Model testing

As in any data-driven learning algorithm, the surrogate models are prone to suffer from overfitting. In other words, the models will tend to perform better on the data they have been trained than on new data that they have not seen before. An example of this is shown in Fig. 3. Therefore, before the surrogate model can be used on parameter estimation samplers, an estimate must be done of its reliability when generating waveforms outside of its own training dataset.

The full waveform dataset from the original Proca catalog is divided into training and testing datasets with 80% and 20% of the samples, respectively. All the model fittings will be performed on the training dataset, and the test set will be reserved as a final evaluation of the model. As a metric for the accuracy of the model we employ the mismatch [83–85] between the predicted and the actual waveform, defined as $1 - M(\psi_1, \psi_2)$. Here, the match $M(\psi_1, \psi_2)$ is given by

$$M(\psi_1, \psi_2) = \max_{t_s, \phi} \frac{\langle \psi_1 | \psi_2 \rangle}{\sqrt{\langle \psi_1 | \psi_1 \rangle \langle \psi_2 | \psi_2 \rangle}}, \quad (3)$$

with $\langle \psi_1 | \psi_2 \rangle$ the inner product defined as

$$\langle \psi_1 | \psi_2 \rangle = 4 \times \text{Re} \int_{\nu_0}^{\nu_1} \frac{\tilde{\psi}_1(\nu) \tilde{\psi}_2^*(\nu)}{S_n(\nu)} d\nu, \quad (4)$$

where $\tilde{\psi}_i(\nu)$ are the Fourier transforms, and $\tilde{\psi}_i^*(\nu)$ their complex conjugates. The match is maximized over relative time shifts t_s and phases ϕ between ψ_1 and ψ_2 . $S_n(f)$ is the one-sided noise power spectral density, which we take to be flat.

D. Model selection

Before the final evaluation on the test set, we have to choose among a number of different architectures of models, as well as the hyperparameters intrinsic to each architecture. It is not convenient, however, to check the performance of each configuration on the final test set. If we did so, we could again be selecting the architecture that best suits our particular test set, without it generalizing to arbitrary data.

While this could be easily solved by splitting the training set into two subsets—one for yet again training and the other for validation—it would further reduce the number of samples in each subset. Given the limited number of simulations in our catalog, it would be convenient to use a scheme where we could extract information from all the data in the original training set. A good solution for this is k -fold cross-validation on the training set. In this scheme, the training set is itself divided in k (approximately) equal parts, called folds. We then train k versions of the model. For each version, we train the model on $k - 1$ of the folds,

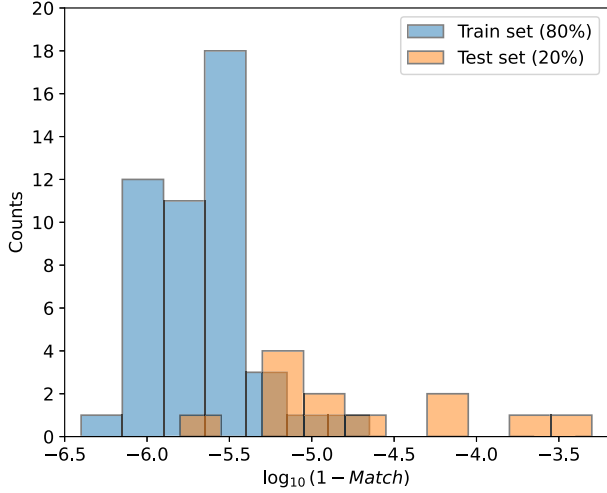


FIG. 3. Histograms of mismatches for the training and testing sets of the SVD surrogate model for the mode $(l, m) = (2, 2)$. Some amount of overfitting can be appreciated, as the model performs worse on the testing data (which has not been used in the training), than on the training data itself.

and we validate it on the remaining fold. With this method, all of the training data samples have been used as a validation of the model architecture, without ever being used both in the training and validation sets. It is important to keep in mind that the final testing set does not intervene at all during the model selection phase. In this paper we will always use $k = 5$.

In Fig. 4 we depict the mismatches of all the $(l, m) = (2, 2)$ waveforms in the dataset, evaluated on their corresponding model, in this case based on the empirical interpolant representation (see Secs. III A and III B). The

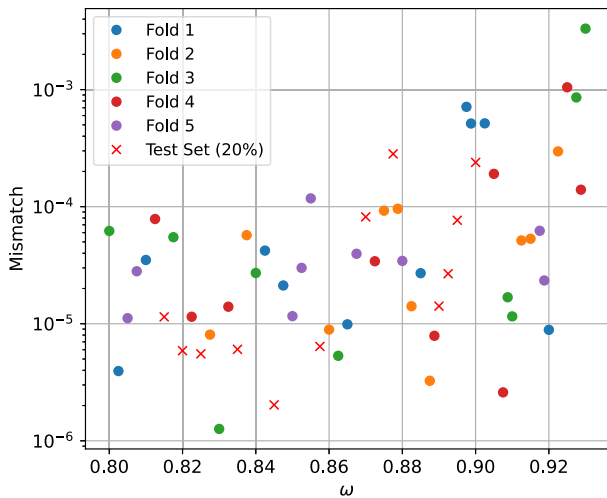


FIG. 4. Mismatches of the empirical interpolant representation model on the dataset, for the mode $(l, m) = (2, 2)$. For each of the five folds in the training set, shown by circular dots, the model has been trained on the other four folds. For the test set, shown by crosses, the model has been trained on the five folds in the training set.

waveforms of the training set, which we depict as circular dots, are divided in the five cross-validation folds, corresponding to the five different colors of the dots. For the validation on each of the folds, the model has been trained on the remaining four folds. On the other hand, for the waveforms in the final test set, the model has been trained on the full training set (all five folds together). The distribution of mismatches is quite similar for all the subsets, which indicates that our statistical estimates of the surrogate performance are consistent with each other. The points at the edge of the domain, close to $\omega = 0.93$, have significantly higher mismatches as extrapolation is being used.

For the sake of consistency, in all of the dataset splittings, we keep the different modes of each simulation together. In other words, if the mode $(l, m) = (2, 2)$ is in one particular subset, then so are the other values of (l, m) .

III. TWO-STAGE MODELS

The two-stage surrogate models are divided in a pipeline of two processes, that we chose to name as encoder-decoder and translator.

- (1) The encoder-decoder phase is an invertible process of dimensional reduction from the N_ψ complex values of ψ_4^i to a feature space of lower dimension N_Z , which can be complex or real depending on the algorithm.
- (2) The translator phase interpolates the features Z^j , as well as the amplitude and peak position, as a function of the physical parameters of the model. In this case, the only physical parameter is the frequency ω/μ_V of the merging Proca stars.

Schematically,

$$\begin{aligned} \mathbb{R} &\rightarrow \mathbb{C}^{N_Z} \rightarrow \mathbb{C}^{N_\psi} \\ \omega &\mapsto Z^j \mapsto \psi_4^i. \end{aligned} \quad (5)$$

When the model is evaluated for parameter estimation, the process is then reversed. First, the physical frequencies are used in the translator to obtain the coordinates Z^j in feature space, which then are passed through the decoder to recover the values of the Newman-Penrose scalar ψ_4 .

A. Singular value decomposition

Our first approach for dimensional reduction is SVD. This produces an orthonormal complex basis for our space of waveforms, in such a way that successive basis vectors are less and less important in the accurate description of the data. The notion of importance is quantitatively encoded in the singular values of the basis. Namely, if we construct an $N_\psi \times N_t$ matrix M , whose N_t columns are the training waveforms, we can decompose it as

$$M = U\Sigma V^\dagger. \quad (6)$$

Here the columns of U are the adapted orthonormal basis u_i , and the diagonal elements of Σ are their corresponding singular values σ_i . By taking the first N_Z of such vectors, we can express a compressed version of any waveform as a linear combination of them. Their N_Z coordinates define the feature space of the model. The fact that U is a unitary matrix greatly simplifies its inversion, as $U^{-1} = U^\dagger$, so the encoding decoding of waveforms becomes trivial.

The main hyperparameter on the SVD models is the number of vectors in the reduced basis N_Z . We have thus performed an analysis of such basis size on the performance of the model, as shown in Fig. 5. The averaged mismatch after the k -fold cross validation process decays approximately as an exponential law for $1 \leq N_Z \leq 7$, and remains fairly constant after that. In other words, having more than seven vectors in the reduced basis does not seem to increase the precision of the model. We have performed our parameter estimation with a basis of 10 vectors, to keep a safe margin.

B. Empirical interpolant representation

The empirical interpolant representation (EIR) is an additional step we take after SVD to obtain a more convenient basis for the space of waveforms in our model. The idea behind EIR algorithms is to choose a reduced basis B_j of the waveform space and a subset of indices

$$\mathcal{I} \subset \mathbb{Z}_{N_\psi}, \quad |\mathcal{I}| = N_Z \quad (7)$$

such that

$$B_j^i = \delta_j^i \quad \forall j \in \mathcal{I}. \quad (8)$$

In other words, the coordinates of $\psi_4(t)$ in the basis B_j are its values at certain selected times $\psi_4(t = T^j) = \psi_4^j$, for $j \in \mathcal{I}$, as

$$\psi_4^i = \sum_{j \in \mathcal{I}} \psi_4^j B_j^i. \quad (9)$$

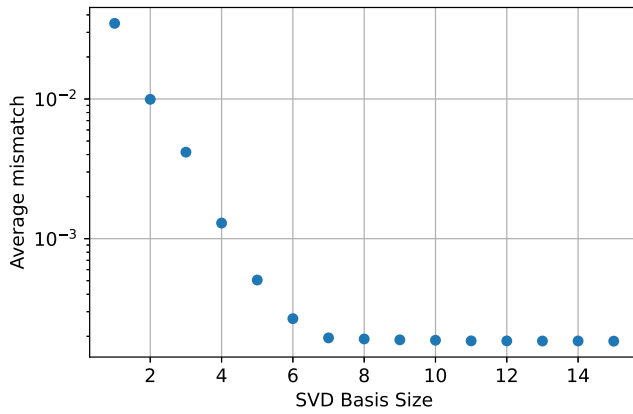


FIG. 5. Average mismatch of the SVD model on the k -fold cross-validation of the test set, for the mode $(l, m) = (2, 2)$. The mismatch decreases exponentially with the number N_Z of vectors until $N_Z = 7$. Adding more vectors does not seem to help.

This process is performed by the PYTHON library ROMPY [75,86], which takes as an input the SVD basis $u_i (i = 1, \dots, N_Z)$ and computes the basis vectors B_j as well as the relevant indices \mathcal{I} .

The encoding procedure of the waveforms into the reduced feature space is then straightforward, as it just involves taking the elements of ψ_4 in the selected positions \mathcal{I} . The decoding is also simple, just multiplying the feature space coordinates by the basis vectors B_j .

In a similar way as in the case of SVD, we have studied the mismatch dependence on the number of EIR basis vectors N_Z , as shown in Fig. 6. The plot looks very similar to that of SVD, with an exponential decay which stabilizes for $N_z \geq 7$.

C. Deep convolutional autoencoder

Autoencoders provide a nonlinear alternative to EIR dimensional reduction. In this case, we set up two 1D convolutional neural networks, which will act as an encoder and a decoder, respectively. The encoder will learn a map from the real and imaginary parts of the waveform ψ_4^i to the feature space Z^j , which is now taken to be real. In other words, the encoder will become a nonlinear mapping from \mathbb{C}^{N_ψ} to \mathbb{R}^{N_z} . On the other hand, the decoder will learn the inverse mapping, trying to recover ψ_4^i from the features Z^j . In the context of autoencoders, the reduced feature space Z^j is typically known as latent space.

The training process is performed on a concatenation of the encoder and decoder networks, in such a way that the goal of the full system is to learn the identity function, with the output being as close as possible to the input. This process is nontrivial as the number of parameters has to be drastically reduced from N_ψ to N_Z as the information is transmitted through the bottleneck between the networks. A schematic representation of this configuration is shown in Fig. 7. We define the loss function of the system as

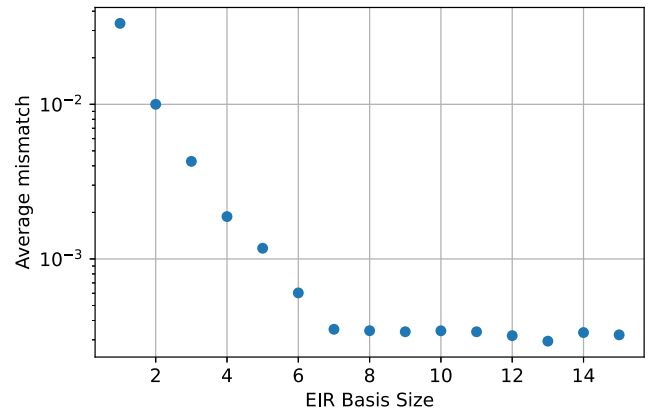


FIG. 6. Average mismatch of the EIR model on the k -fold cross-validation of the test set, for the mode $(l, m) = (2, 2)$. The mismatch decreases exponentially with the number N_Z of vectors until $N_Z = 7$. Adding more vectors does not seem to help.

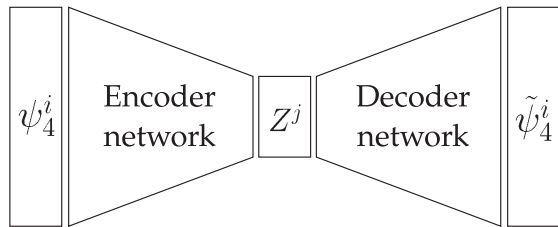


FIG. 7. Schematic diagram of the autoencoder architecture. Two deep convolutional networks, the encoder and the decoder, have to reconstruct the original signal by encoding the information in a reduced latent space. Once the autoencoder is trained, the decoder can be used on its own to produce waveforms from vectors of coordinates Z^j in the latent space.

$$\mathcal{L} = \frac{1}{2N_\psi} \sum_{i=1}^{N_\psi} |\psi_4^i - \tilde{\psi}_4^i|^2, \quad (10)$$

with ψ_4^i and $\tilde{\psi}_4^i$ the original and recovered waveforms, respectively. The loss is also averaged over the instances in the training batch. After the training, the encoder and decoder modules can be used separately to map the waveforms to their coordinates in the latent space and vice versa. Figure 8 shows the embedding of the waveforms in a latent space of $N_Z = 2$, where the model has spontaneously ordered the waveforms by their value of ω/μ_V . In this case the mapping is nonlinear, so the number of coordinates in the latent space is usually lower than the number of basis vectors that are needed in SVD or EIR encoding.

The structure of the networks is fairly symmetric between encoder decoder. The encoder receives the training minibatch of waveforms as a tensor of dimensions

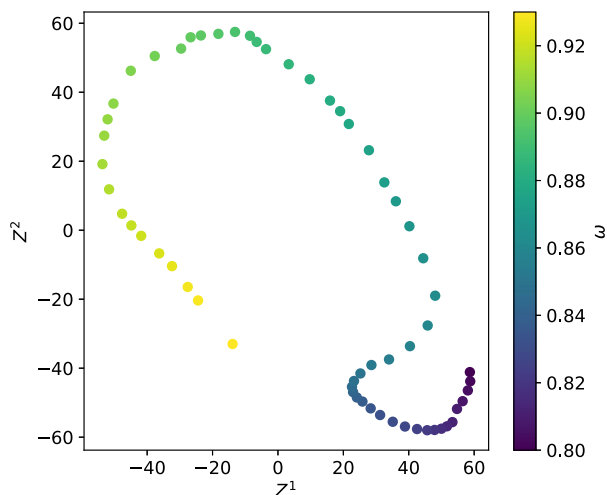


FIG. 8. Distribution of the dataset waveforms in the two-dimensional latent space, with coordinates (Z^1, Z^2) , for the mode $(l, m) = (2, 2)$. The autoencoder has autonomously ordered the waveforms by their governing physical parameter ω/μ_V along a continuous curve embedded in the latent space.

$(N_B, 2, N_\psi)$. $N_B = 16$ is the number of instances in the training minibatch, two is the number of channels (corresponding to the real and imaginary parts of the wave), and $N_\psi = 512$ is the length of the time series containing the wave. The batch is first passed through a one-dimensional convolutional layer of 16 filters of kernel size 4, stride 2, and padding 1. After that a leaky ReLU activation function, with a negative slope of 0.2, is applied. Then the batch goes through a series of six more convolutions with identical kernel configurations, but each time duplicating the number of filters. After each convolution, we apply a batch normalization and another leaky ReLU with a negative slope of 0.2. This configuration of filters reduces the length of the data samples by 2 at each convolution, while doubling the number of channels. Finally a last convolution is applied, this time with a stride of 1 and padding reduced to 0, with N_Z filters. This reduces the minibatch to a tensor of size $(N_B, N_Z, 1)$, where the N_Z channel values are taken to be the coordinates in the latent space.

The architecture of the decoder is quite exactly inverse from the encoder. The input minibatch of dimensions $(N_B, N_Z, 1)$ is passed through a transposed convolution with 64 filters of kernel size 4, stride 1, and padding 0. Then, a series of 7 more (transposed) convolutions are applied, now with a stride of 2 and padding of 1, each one duplicating the size of the data instances and halving the number of filters. After each convolution (except the very last one), we apply batch normalization and a ReLU activation. The last convolution actually reduces the number of filters down to 2 (real and imaginary parts of ψ_4), resulting in an output of dimensions $(N_B, 2, N_\psi)$, as expected.

The training is performed on a Nvidia RTX 4090 24GB GPU over minibatches of 16 waveforms going through 500 iteration epochs with a learning rate of 10^{-3} , using an Adam optimizer with parameters $(\beta_1, \beta_2) = (0.5, 0.999)$, in the framework PYTORCH [87].

Figure 9 shows the average mismatch of the surrogate as a function of the latent space dimension N_Z , computed over the k -fold cross validation process. Even though the waveforms depend on a single physical parameter, namely the frequency ω/μ_V , the model has proven to perform much better if we allow at least two dimensions in the latent space. A latent dimension higher than that does not have a large influence on the mismatch. For this reason, the final selected model has been chosen to have $N_Z = 2$.

D. The translator phase

The task of the translator phase of the model is to interpolate the coordinates in feature space F^j as a function of the Proca star frequency ω/μ_V . As there is only one independent variable, order 3 spline interpolation is an efficient and fast method. These are provided by the `scipy.interpolate` function `interp1d`, which is applied to every coordinate in the latent space, as well

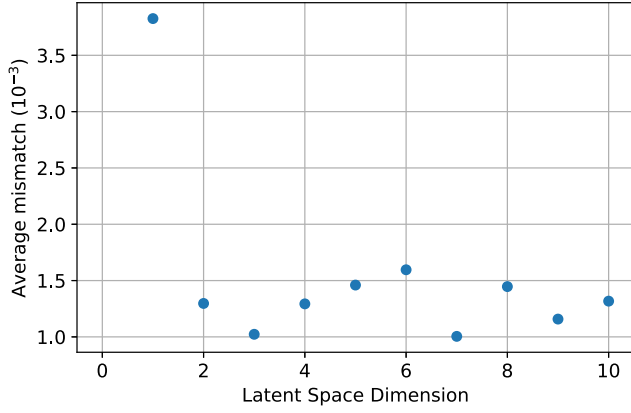


FIG. 9. Average mismatch of the autoencoder model on the k -fold cross-validation of the test set, for the mode $(l, m) = (2, 2)$. The performance does not seem to vary as long as the latent space dimension N_Z is strictly larger than 1.

as the peak position and amplitudes (t_0, A) , as a function of the variable ω/μ_V .

IV. MONOLITHIC MODEL

We have, for completeness, trained a monolithic model based on fully connected neural networks. In this case, instead of performing a preliminary dimensional reduction before the final interpolation in ω/μ_V , we allow a single neural network to receive ω/μ_V and produce ψ_4^i directly. In this case, the real and imaginary part of the waveform are concatenated in a single vector of dimension $2N_\psi = 1024$, as complex values are not trivially handled by common activation functions. Cubic splines are used to learn the dependence of the peak position and amplitudes (t_0, A) , as defined in Sec. II B.

The wave-fitting network has an input layer with 1 neuron (receiving ω/μ_V) and successive layers of sizes 64,

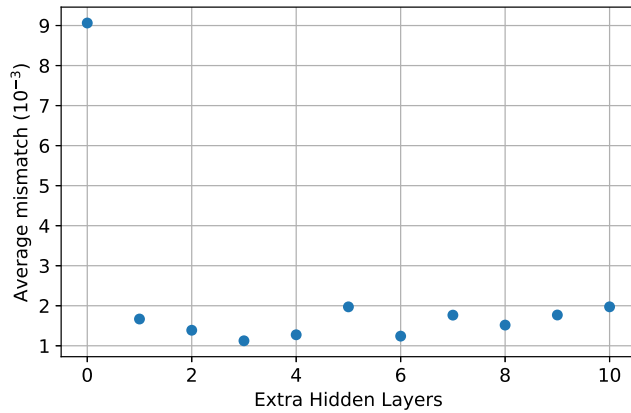


FIG. 10. Average mismatch of the monolithic model on the k -fold cross-validation of the test set, for the mode $(l, m) = (2, 2)$. The performance does not seem to vary as long as we include at least one extra hidden layer of 64 neurons.

256, 512, and 1024 (the output layer). We use GELU activation functions after each layer except the last one.

As in the case of the autoencoder, training is performed on a Nvidia RTX 4090 24GB GPU over 500 iteration epochs divided in minibatches of 16 waveforms, with an Adam optimizer of learning rate of 10^{-3} and $(\beta_1, \beta_2) = (0.5, 0.999)$.

In order to gauge the dependence of the model on hyperparameters, we introduce a number of replicas of the layer of size 64, exclusively on the wave-fitting network. As an example, the network with two extra hidden layers would have sizes $(1, 64, 64, 64, 256, 512, 1024)$. In Fig. 10 we observe that the results improve significantly when we

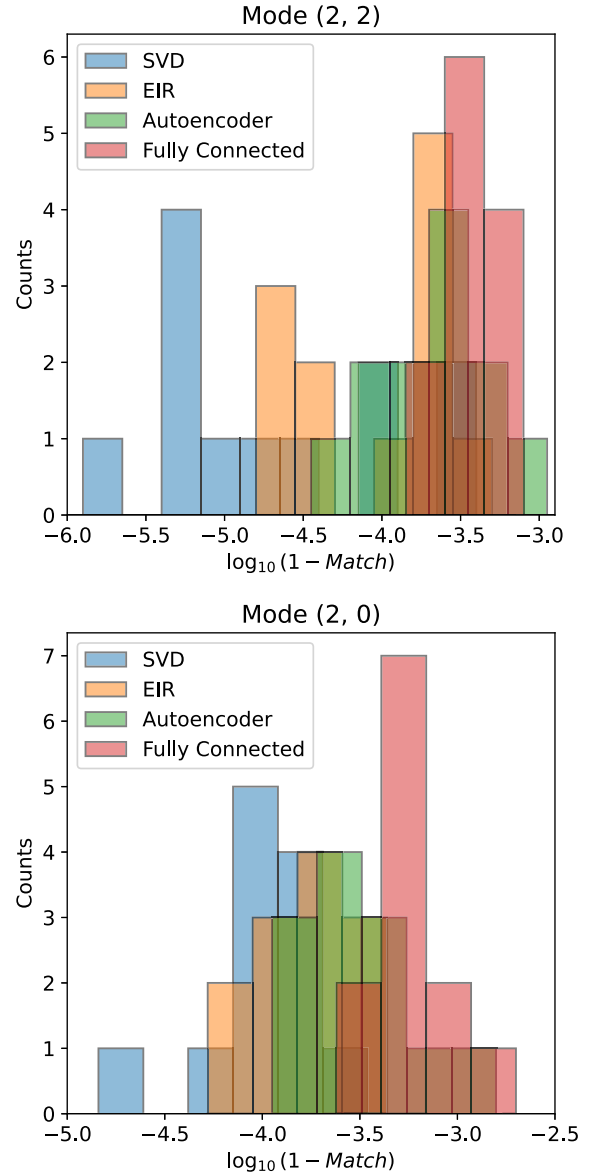


FIG. 11. Histograms of the distribution of the mismatches on the final test set for the four tested model architectures. Lower mismatches correspond to a better model accuracy. We consider the two dominant modes: $(l, m) = (2, 2)$ in the upper plot and $(l, m) = (2, 0)$ in the lower plot.

add one extra layer but plateau after that. For this reason, we used one extra layer for the final model.

V. RESULTS

A. Model evaluation

Figure 11 shows histograms of the mismatches obtained on the final testing set by the four categories of models. We consider the two dominant modes, namely $(l, m) = (2, 2)$ and $(l, m) = (2, 0)$. In both cases we can observe that the SVD model seems to work best, followed by EIR, the autoencoder, and finally the monolithic fully connected network. There is, however, a significant overlap between the distributions.

There are several possible reasons for the better performance of SVD. A lower number of parameters makes the model less prone to overfitting and therefore more robust. Also, SVD is based on standard linear algebra algorithms, which do not require training by gradient descent, which removes the possible accuracy losses due to early or late stopping of the training process.

The significantly better performance of the SVD model on the testing set, together with its simplicity and the fact that its training is much faster than those based on neural networks, clearly indicate that this type of model is the most convenient for our physical system. Indeed, the parameter estimation Bayesian parameter inferences in

Secs. VB and VC have been performed with SVD models with 10 basis vectors and $N_\psi = 3001$, in this case trained with the complete dataset of 59 waveforms.

B. Injection recovery

We test the performance of our model in parameter inference and model selection tasks. To this end, we inject in zero noise a family of 30 signals from head-on Proca-star mergers of varying ω/μ_V , directly obtained through numerical-relativity simulations. We recover these injections using our SVD surrogate model together with the parameter inference code BILBY [15] in its parallelizable version PARALLEL BILBY [88]. We set uniform priors in the detector-frame total mass and ω/μ_V . In addition, we set isotropic priors both on sky location and source orientation, together with a uniform prior in the polarization angle and a prior in distance uniform in comoving volume. These priors match those used for the analyses presented in [43,45] restricted equal-mass systems. We choose a three-detector network formed by the two Advanced LIGO detectors and Advanced Virgo equipped with their power spectral density at the time of the GW190521 event. We perform our analysis directly on the Newman-Penrose scalar ψ_4 , as described in [43]. We sample the parameter space using the sampler DYNESTY [89] with $N = 1024$ live points.

Figure 12 shows the posterior distributions, together with 90% credible intervals, for the field frequency ω/μ_V , as a

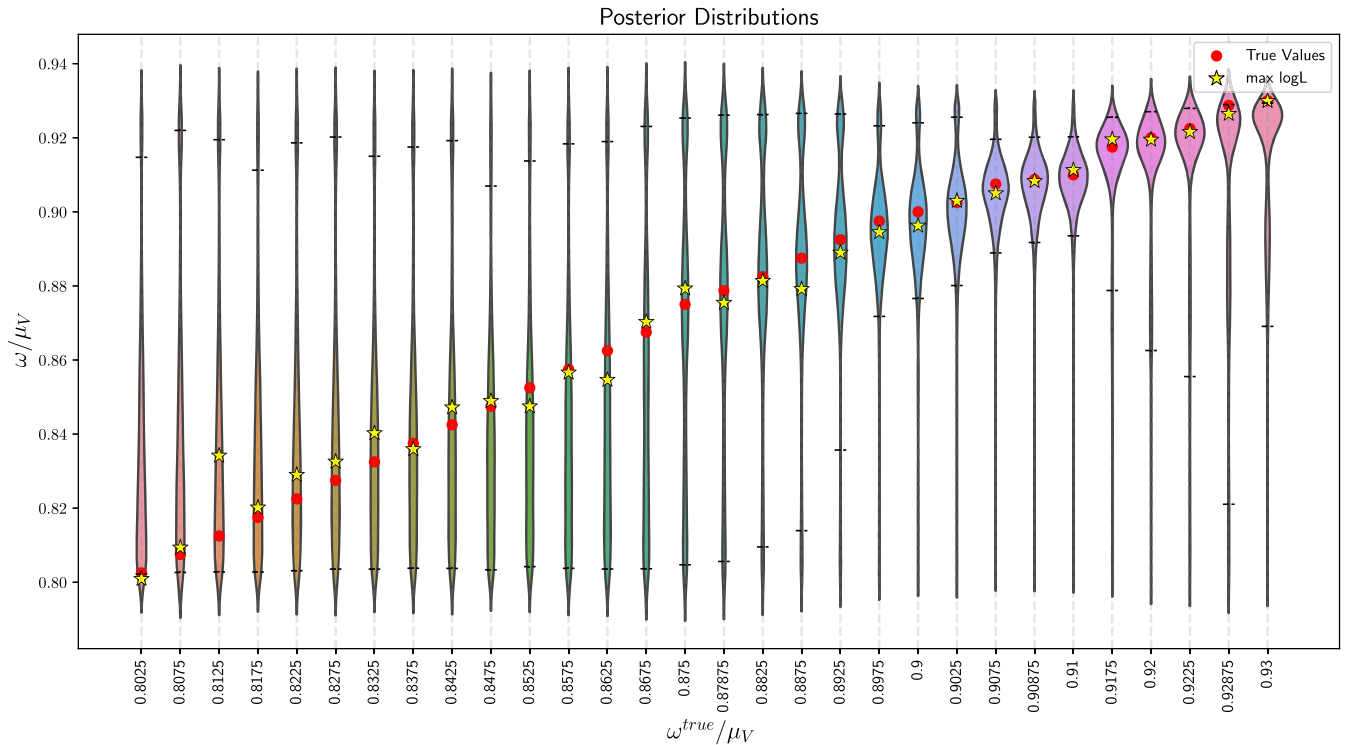


FIG. 12. Posterior distributions for the field frequency ω/μ_V as a function of the corresponding true values. We obtain these through the injection of numerical relativity waveforms in zero noise and their posterior recovery using our surrogate model. True values are denoted by red dots (red line) in the lower (upper) plot while maximum likelihood values are denoted by yellow stars. The horizontal bars delimit 90% credible intervals.

function of the true injected value $\omega_{\text{true}}/\mu_V$. The true values are denoted by red circles while maximum likelihood (ML) values (i.e., those producing the best fits) are denoted by yellow stars. As somewhat expected from our match

analyses, we find that the posterior distribution always peaks very close to the true value, with the latter being always contained within the 90% credible intervals. The same is found in most cases for the ML values, which

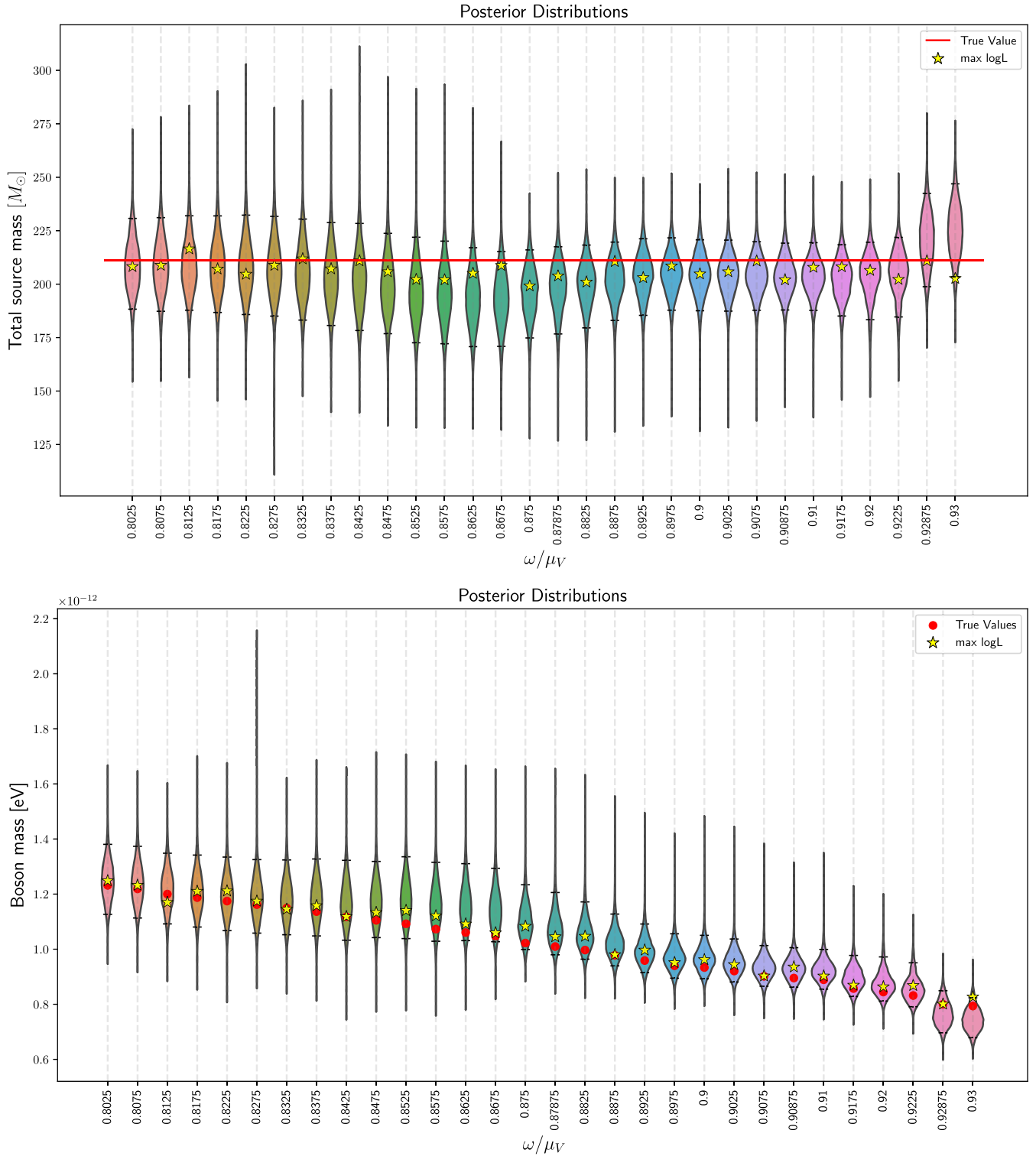


FIG. 13. Posterior distributions for the source-frame total mass (top) and the ultralight boson mass (bottom) as a function of the value of ω/μ_V . True values are denoted by red dots in the plot while maximum likelihood values are denoted by yellow stars. The horizontal bars delimit 90% credible intervals.

always lay very close to the true parameters found by the sampler. We note that the sampler is not designed to find the actual best-fit parameters but to reconstruct the corresponding posterior probability distributions. Due to this, in some cases, the recovered ML parameters may not match the true ones, even if the waveform model is infinitely accurate. This is the case, for instance, when there are large degeneracies in the parameter space or when the sampling is not extremely aggressive. Nevertheless, since in certain situations parameter recovery can be strongly influenced by Bayesian priors causing shifts in the posteriors, we report the location of the ML points to show that such shifts are indeed sourced by priors and not by systematic errors due to the waveform model.

We find that certain values of ω_{true} lead to varying features in the posteriors. First, some cases lead to wide distributions while others lead to significantly sharper ones. This reflects the fact that, in certain regions of the parameter space, small changes in ω_{true} cause significant changes in the waveform morphology while smoother changes are induced in other regions. Second, for some cases we find somewhat bimodal distributions. Far from denoting any issues in the analysis or systematic errors in the waveform model, these features come as a consequence of the physics governing Proca-star waveforms. In particular, mergers corresponding to the lower and upper end of our ω/μ_V range, are significantly louder than those corresponding to intermediate frequencies. Therefore, such configurations are highly favored by our distance prior, producing secondary peaks of the posterior distribution either at large frequencies (for the lowest frequency injections), both large and low frequencies (for the midfrequency injections), and almost imperceptible peaks at low frequencies for the highest-frequency injections. For the latter, the posterior also somewhat extends artificially to lower frequencies due to the upper frequency cutoff imposed by the model, which prevents the posterior distribution to extend beyond such cutoff.

Figure 13 shows the posterior distributions for the total mass M_{total} of the binary and the corresponding boson mass μ_V . First, we note that the true values are again always contained within the 90% credible intervals. Second, we note that the posteriors for M_{total} are systematically shifted towards lower values. Again, far from reflecting systematic errors in our model, this is due to the impact of the prior for the luminosity distance [43,45]. Recall that the source-frame total mass is $M_{\text{total}} = M_{\text{det}}/(1+z)$, where M_{det} and z are, respectively, the detector-frame total mass and the redshift. Since the distance prior strongly favors large distances, we find that such estimates, and therefore those for z , are always shifted towards large values, causing the opposite effect on M_{total} . The same effect is present but less pronounced for μ_V , owing to its strong dependency on the—better captured—value of ω/μ_V . Finally, we note that ML values are always closer (when not directly on top) to

the true injected values than the bulk of the posterior distributions.

C. GW190521 parameter estimation

We perform parameter inference on the GW event GW190521, which has been previously analyzed directly using numerical relativity simulations for head-on Proca-star waveforms, both restricted to equal-mass cases [43,45] and extended to unequal-mass ones [44]. We set the exact same priors described in the previous section, which match those used in [43,45]. We will compare our results to those reported in the second column of Table III in [43], which unlike those in [45], make direct use of the Newman-Penrose scalar. The two main differences between these two analyses are (i) the usage of a surrogate model instead of numerical relativity waveforms and (ii) the continuous sampling of the parameter space as opposed to the exploration of a discrete set of values of ω/μ_V .

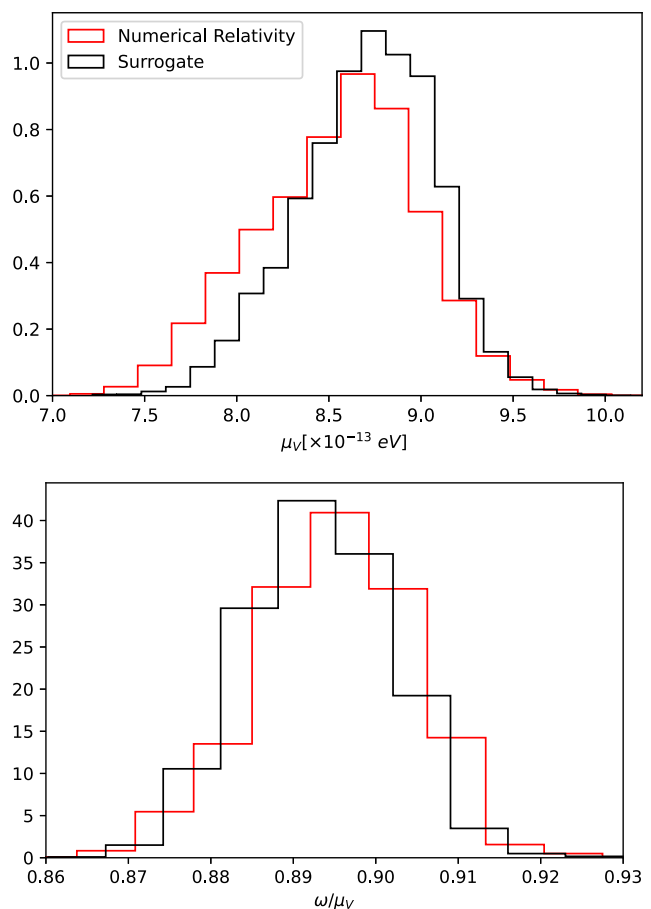


FIG. 14. Posterior probability distributions for the field frequency ω/μ_V and the ultralight boson mass μ_V for GW190521. The black curve corresponds to the analysis carried out with our surrogate model. The red curve corresponds to the analysis carried out directly with numerical relativity templates, using the exact framework described in [43,44].

TABLE I. Parameters for GW190521 obtained under the head-on Proca-star merger analysis using our surrogate model (middle column) and the approach based on numerical relativity waveforms presented in [43] (right column). We quote median values with symmetric 90% credible intervals. The last two rows reports the natural logarithmic Bayes factor for the signal vs noise hypotheses and the corresponding maximum likelihoods.

Parameter	Surrogate	NR
Total/final mass [M_{\odot}]	233^{+10}_{-12}	233^{+12}_{-16}
Inclination [rad]	$0.84^{+0.21}_{-0.34}$	$0.85^{+0.23}_{-0.36}$
Luminosity distance [Mpc]	549^{+240}_{-149}	544^{+296}_{-163}
Redshift z	$0.11^{+0.03}_{-0.05}$	$0.11^{+0.03}_{-0.05}$
Total red-shifted mass [M_{\odot}]	259^{+7}_{-8}	260^{+8}_{-9}
Field frequency ω/μ_V	$0.893^{+0.014}_{-0.014}$	$0.890^{+0.015}_{-0.015}$
Boson mass μ_V [$\times 10^{-13}$ eV]	$8.73^{+0.53}_{-0.68}$	$8.60^{+0.63}_{-0.62}$
ln Bayes factor S/N	90.46	90.36
ln max. likelihood	116.1	116.5

Table I reports the parameter estimates obtained from our analysis together with those obtained in [43], expressed as median values together with symmetric 90% credible intervals. In addition, Fig. 14 shows the corresponding posterior distributions for the field frequency ω/μ_V and the ultralight-boson mass μ_V . We note that the results are in very good agreement, indicating again the reliability of our surrogate model. Table I also shows the corresponding natural log-Bayes’ factors with respect to the noise hypothesis, which match within the sampling uncertainty of 0.1. Finally, we find that the numerical relativity model recovers a slightly larger log likelihood, indicating that the best-fitting waveform fits the data marginally better than in the surrogate case. We attribute this to the following fact. While for the case of the surrogate model we perform a single inference run over all waveform parameters, for the NR model we combine the result of many runs that sample over the extrinsic waveform parameters and total mass for fixed ω/μ_V [45], which leads to a more aggressive sampling of the parameter space.

VI. DISCUSSION

We have constructed several accurate surrogate models that generate waveforms of head-on collisions of equal-mass Proca stars. The models are purely data driven, without any predefined physical input, and are trained on a catalog of 59 numerical relativity waveforms with several multipoles each. Different model architectures have been analyzed. On the one hand, a collection of two-stage models, which implement a dimensional reduction algorithm followed by cubic spline interpolation of the governing parameters. The dimensional reduction algorithms discussed comprise singular value decomposition, empirical interpolant representation, and finally a deep convolutional autoencoder. On the other hand,

we have also tested a monolithic model where a fully connected neural network interpolates the full waveform time series as a function of the physical parameter ω/μ_V .

While all four models have proven to be accurate, with all mismatches below 10^{-3} for the mode $(l, m) = (2, 2)$, the model based on singular value decomposition has proven to be the most precise. In terms of precision, SVD is followed by the empirical interpolant representation, the autoencoder and finally the monolithic fully connected network. The two first models are not only the most precise but also much faster than those based on neural networks, as they do not require gradient descent optimization. However, the neural networks are able to encode waveforms nonlinearly from very few input real parameters (2 for the autoencoder, 1 for the fully connected), while SVD and EIR require at least seven complex parameters for our dataset.

We have tested the SVD surrogate model within the task of Bayesian parameter inference through the recovery of numerical relativity injections in zero noise, obtaining excellent results. Finally, we have also performed parameter estimation on GW190521 data, finding strong agreement with the results obtained directly using numerical relativity templates. As a general summary, we conclude that the type of surrogate models investigated in this work can be regarded as a convenient tool to efficiently perform Bayesian analysis on gravitational waveforms from bosonic star mergers.

A natural extension of this work is the inclusion of waveforms from Proca star mergers with unequal masses. Our early attempts at this extension, however, have encountered serious difficulties due to the interference effects between the stars at merger [42]. When the Proca fields from the two colliding stars oscillate at different frequencies ω_1 and ω_2 , the relative phase at the point of encounter depends on the particular values of these frequencies. This causes an interference pattern, where the amplitude of the emitted gravitational waves presents strong variations. Additionally, at the regions of lowest amplitude, the shape of the waveform changes very rapidly with ω_1 and ω_2 . At present, our catalog of simulations is not sufficiently dense in these difficult regions for the model to reach the desired precision.

ACKNOWLEDGMENTS

We would like to thank Koustav Chandra for his very useful comments on the first manuscript. R.L. acknowledges financial support provided by Generalitat Valenciana/Fons Social Europeu through APOSTD 2022 postdoctoral Grant No. CIAPOS/2021/150. J.C.B. is supported by a fellowship from “la Caixa” Foundation (ID 100010434) and from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 847648.

The fellowship code is LCF/BQ/PI20/11760016. J. C. B. is also supported by the research Grant No. PID2020–118635GB-I00 from the Spain-Ministerio de Ciencia e Innovación. N. S. G. acknowledges support from the Spanish Ministry of Science and Innovation via the Ramón y Cajal programme (Grant No. RYC2022-037424-I), funded by Grant No. MCIN/AEI/10.13039/501100011033 and by “ESF Investing in your future.” M. L.-M., N. S. G., A. T. F., and J. A. F. are supported by the Spanish Agencia Estatal de Investigación (Grant No. PID2021-125485NB-C21) funded by Grant No. MCIN/AEI/10.13039/501100011033 and ERDF A way of making Europe, and the Generalitat Valenciana (Grant No. CIPROM/2022/49). We acknowledge further support by the European Horizon Europe staff exchange (SE) programme HORIZON-MSCA2021-SE-01 Grant No. NewFunFiCO-101086251. This work is supported by the Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT—Fundação para a Ciência e a Tecnologia), reference No. UIDB/04106/2020 and No. UIDP/04106/2020 (Refs. [90,91]). The authors acknowledge support from the projects No. PTDC/FIS-AST/3041/2020, No. CERN/FIS-PAR/0024/2021 and No. 2022.04560.PTDC (Refs. [92–94]). The authors acknowledge computational resources provided by the CIT cluster of the LIGO Laboratory and supported by National Science Foundation Grants No. PHY-0757058 and No. PHY0823459; and the support of the NSF CIT cluster for the provision of computational resources for our parameter inference runs. This material is based upon

work supported by NSF’s LIGO Laboratory which is a major facility fully funded by the National Science Foundation. The analyzed LIGO-Virgo data and the corresponding power spectral densities, in their strain versions, are publicly available at the online Gravitational-Wave Open Science Center [95,96]. This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gwosc.org), a service of LIGO Laboratory, the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. LIGO Laboratory and Advanced LIGO are funded by the United States National Science Foundation (NSF) as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. K. A. G. R. A. is supported by Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan. This manuscript has LIGO DCC No. P2400107.

-
- [1] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), Observation of gravitational waves from a binary black hole merger, *Phys. Rev. Lett.* **116**, 061102 (2016).
 - [2] 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).
 - [3] B. P. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW170814: A three-detector observation of gravitational waves from a binary black hole coalescence, *Phys. Rev. Lett.* **119**, 141101 (2017).
 - [4] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW190521: A binary black hole merger with a total mass of $150M_{\odot}$, *Phys. Rev. Lett.* **125**, 101102 (2020).
 - [5] 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).
 - [6] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GWTC-2: Compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run, *Phys. Rev. X* **11**, 021053 (2021).
 - [7] R. Abbott *et al.* (KAGRA, VIRGO, and LIGO Scientific Collaborations), GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, *Phys. Rev. X* **13**, 041039 (2023).
 - [8] S. Olsen, T. Venumadhav, J. Mushkin, J. Roulet, B. Zackay, and M. Zaldarriaga, New binary black hole mergers in the LIGO-Virgo O3a data, *Phys. Rev. D* **106**, 043009 (2022).
 - [9] A. H. Nitz, S. Kumar, Y.-F. Wang, S. Kastha, S. Wu, M. Schäfer, R. Dhurkunde, and C. D. Capano, 4-OGC: Catalog of gravitational waves from compact binary mergers, *Astrophys. J.* **946**, 59 (2023).
 - [10] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries, *Phys. Rev. D* **85**, 122006 (2012).

- [11] S. A. Usman *et al.*, The PyCBC search for gravitational waves from compact binary coalescence, *Classical Quantum Gravity* **33**, 215004 (2016).
- [12] K. Cannon *et al.*, Gstlal: A software framework for gravitational wave discovery, *SoftwareX* **14**, 100680 (2021).
- [13] Q. Chu, M. Kovalam, L. Wen, T. Slaven-Blair, J. Bosveld, Y. Chen, P. Clearwater, A. Codoreanu, Z. Du, X. Guo, X. Guo, K. Kim, T. G. F. Li, V. Oloworaran, F. Panther, J. Powell, A. S. Sengupta, K. Wette, and X. Zhu, SPIIR online coherent pipeline to search for gravitational waves from compact binary coalescences, *Phys. Rev. D* **105**, 024023 (2022).
- [14] 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).
- [15] G. Ashton *et al.*, BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy, *Astrophys. J. Suppl. Ser.* **241**, 27 (2019).
- [16] B. Zackay, L. Dai, T. Venumadhav, J. Roulet, and M. Zaldarriaga, Detecting gravitational waves with disparate detector responses: Two new binary black hole mergers, *Phys. Rev. D* **104**, 063030 (2021).
- [17] K. Chandra, V. Villa-Ortega, T. Dent, C. McIsaac, A. Pai, I. W. Harry, G. S. C. Davies, and K. Soni, An optimized PyCBC search for gravitational waves from intermediate-mass black hole mergers, *Phys. Rev. D* **104**, 042004 (2021).
- [18] K. Chandra, J. Calderón Bustillo, A. Pai, and I. W. Harry, First gravitational-wave search for intermediate-mass black hole mergers with higher-order harmonics, *Phys. Rev. D* **106**, 123003 (2022).
- [19] S. Babak, R. Balasubramanian, D. Churches, T. Cokelaer, and B. S. Sathyaprakash, A template bank to search for gravitational waves from inspiralling compact binaries: I. Physical models, *Classical Quantum Gravity* **23**, 5477 (2006).
- [20] S. Sakon *et al.*, Template bank for compact binary mergers in the fourth observing run of Advanced LIGO, Advanced VIRGO, and KAGRA, *Phys. Rev. D* **109**, 044066 (2024).
- [21] V. Cardoso and P. Pani, Testing the nature of dark compact objects: A status report, *Living Rev. Relativity* **22**, 4 (2019).
- [22] D. J. Kaup, Klein-Gordon geon, *Phys. Rev.* **172**, 1331 (1968).
- [23] R. Ruffini and S. Bonazzola, Systems of self-gravitating particles in general relativity and the concept of an equation of state, *Phys. Rev.* **187**, 1767 (1969).
- [24] A. Arvanitaki, S. Dimopoulos, S. Dubovsky, N. Kaloper, and J. March-Russell, String axiverse, *Phys. Rev. D* **81**, 123530 (2010).
- [25] A. Arvanitaki and S. Dubovsky, Exploring the string axiverse with precision black hole physics, *Phys. Rev. D* **83**, 044026 (2011).
- [26] F. F. Freitas, C. A. R. Herdeiro, A. P. Morais, A. Onofre, R. Pasechnik, E. Radu, N. Sanchis-Gual, and R. Santos, Ultralight bosons for strong gravity applications from simple standard model extensions, *J. Cosmol. Astropart. Phys.* **12** (2021) 047.
- [27] C. A. R. Herdeiro, J. Kunz, I. Perapechka, E. Radu, and Y. Shnir, Multipolar boson stars: Macroscopic Bose-Einstein condensates akin to hydrogen orbitals, *Phys. Lett. B* **812**, 136027 (2021).
- [28] F. E. Schunck and E. W. Mielke, Rotating boson stars, in *Relativity and Scientific Computing: Computer Algebra, Numerics, Visualization* (Springer, New York, 1996), pp. 138–151.
- [29] C. Herdeiro, I. Perapechka, E. Radu, and Y. Shnir, Asymptotically flat spinning scalar, Dirac and Proca stars, *Phys. Lett. B* **797**, 134845 (2019).
- [30] R. Brito, V. Cardoso, C. A. R. Herdeiro, and E. Radu, Proca stars: Gravitating Bose–Einstein condensates of massive spin 1 particles, *Phys. Lett. B* **752**, 291 (2016).
- [31] N. Sanchis-Gual, F. Di Giovanni, M. Zilhão, C. Herdeiro, P. Cerdá-Durán, J. A. Font, and E. Radu, Nonlinear dynamics of spinning bosonic stars: Formation and stability, *Phys. Rev. Lett.* **123**, 221101 (2019).
- [32] F. Di Giovanni, N. Sanchis-Gual, P. Cerdá-Durán, M. Zilhão, C. Herdeiro, J. A. Font, and E. Radu, Dynamical bar-mode instability in spinning bosonic stars, *Phys. Rev. D* **102**, 124009 (2020).
- [33] N. Siemonsen and W. E. East, Stability of rotating scalar boson stars with nonlinear interactions, *Phys. Rev. D* **103**, 044022 (2021).
- [34] A. S. Dmitriev, D. G. Levkov, A. G. Panin, E. K. Pushnaya, and I. I. Tkachev, Instability of rotating Bose stars, *Phys. Rev. D* **104**, 023504 (2021).
- [35] N. Sanchis-Gual, F. Di Giovanni, C. Herdeiro, E. Radu, and J. A. Font, Multifield, multifrequency bosonic stars and a stabilization mechanism, *Phys. Rev. Lett.* **126**, 241105 (2021).
- [36] E. Seidel and W.-M. Suen, Formation of solitonic stars through gravitational cooling, *Phys. Rev. Lett.* **72**, 2516 (1994).
- [37] F. Di Giovanni, N. Sanchis-Gual, C. A. R. Herdeiro, and J. A. Font, Dynamical formation of Proca stars and quasistationary solitonic objects, *Phys. Rev. D* **98**, 064044 (2018).
- [38] C. A. R. Herdeiro, E. Radu, N. Sanchis-Gual, N. M. Santos, and E. dos Santos Costa Filho, The non-spherical ground state of Proca stars, *Phys. Lett. B* **852**, 138595 (2024).
- [39] S. L. Liebling and C. Palenzuela, Dynamical boson stars, *Living Rev. Relativity* **26**, 1 (2023).
- [40] N. Sanchis-Gual, C. Herdeiro, E. Radu, J. C. Degollado, and J. A. Font, Numerical evolutions of spherical Proca stars, *Phys. Rev. D* **95**, 104028 (2017).
- [41] N. Sanchis-Gual, C. Herdeiro, J. A. Font, E. Radu, and F. Di Giovanni, Head-on collisions and orbital mergers of Proca stars, *Phys. Rev. D* **99**, 024017 (2019).
- [42] N. Sanchis-Gual, J. Calderón Bustillo, C. Herdeiro, E. Radu, J. A. Font, S. H. W. Leong, and A. Torres-Forné, Impact of the wavelike nature of Proca stars on their gravitational-wave emission, *Phys. Rev. D* **106**, 124011 (2022).
- [43] J. Calderon Bustillo, I. C. F. Wong, N. Sanchis-Gual, S. H. W. Leong, A. Torres-Forne, K. Chandra, J. A. Font, C. Herdeiro, E. Radu, and T. G. F. Li, Gravitational-wave parameter inference with the Newman-Penrose scalar, *Phys. Rev. X* **13**, 041048 (2023).
- [44] J. Calderon Bustillo, N. Sanchis-Gual, S. H. W. Leong, K. Chandra, A. Torres-Forne, J. A. Font, C. Herdeiro, E. Radu, I. C. F. Wong, and T. G. F. Li, Searching for vector boson-star mergers within LIGO-Virgo intermediate-mass black-hole merger candidates, *Phys. Rev. D* **108**, 123020 (2023).

- [45] J. Calderón Bustillo, N. Sanchis-Gual, A. Torres-Forné, J. A. Font, A. Vajpeyi, R. Smith, C. Herdeiro, E. Radu, and S. H. W. Leong, GW190521 as a merger of Proca stars: A potential new vector boson of 8.7×10^{-13} eV, *Phys. Rev. Lett.* **126**, 081101 (2021).
- [46] P. Schmidt, Gravitational waves from binary black hole mergers: Modeling and observations, *Front. Astron. Space Sci.* **7**, 28 (2020).
- [47] A. Buonanno and T. Damour, Effective one-body approach to general relativistic two-body dynamics, *Phys. Rev. D* **59**, 084006 (1999).
- [48] A. Buonanno and T. Damour, Transition from inspiral to plunge in binary black hole coalescences, *Phys. Rev. D* **62**, 064015 (2000).
- [49] A. Bohe *et al.*, 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).
- [50] S. Ossokine, A. Buonanno, S. Marsat, R. Cotesta, S. Babak, T. Dietrich, R. Haas, I. Hinder, H. P. Pfeiffer, M. Pürrer, C. J. Woodford, M. Boyle, L. E. Kidder, M. A. Scheel, and B. Szilágyi, Multipolar effective-one-body waveforms for precessing binary black holes: Construction and validation, *Phys. Rev. D* **102**, 044055 (2020).
- [51] 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).
- [52] A. Nagar, P. Retegno, R. Gamba, S. Albanesi, A. Albertini, and S. Bernuzzi, Analytic systematics in next generation of effective-one-body gravitational waveform models for future observations, *Phys. Rev. D* **108**, 124018 (2023).
- [53] A. Gonzalez, R. Gamba, M. Breschi, F. Zappa, G. Carullo, S. Bernuzzi, and A. Nagar, Numerical-relativity-informed effective-one-body model for black-hole–neutron-star mergers with higher modes and spin precession, *Phys. Rev. D* **107**, 084026 (2023).
- [54] P. Ajith *et al.*, Phenomenological template family for black-hole coalescence waveforms, *Classical Quantum Gravity* **24**, S689 (2007).
- [55] L. Santamaria *et al.*, Matching post-Newtonian and numerical relativity waveforms: Systematic errors and a new phenomenological model for non-precessing black hole binaries, *Phys. Rev. D* **82**, 064016 (2010).
- [56] S. Khan, F. Ohme, K. Chatziioannou, and M. Hannam, Including higher order multipoles in gravitational-wave models for precessing binary black holes, *Phys. Rev. D* **101**, 024056 (2020).
- [57] G. Pratten, C. García-Quirós, M. Colleoni, A. Ramos-Buades, H. Estellés, M. Mateu-Lucena, R. Jaume, M. Haney, D. Keitel, J. E. Thompson, and S. Husa, Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes, *Phys. Rev. D* **103**, 104056 (2021).
- [58] 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).
- [59] J. Blackman, S. E. Field, C. R. Galley, B. Szilágyi, M. A. Scheel, M. Tiglio, and D. A. Hemberger, Fast and accurate prediction of numerical relativity waveforms from binary black hole coalescences using surrogate models, *Phys. Rev. Lett.* **115**, 121102 (2015).
- [60] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, D. A. Hemberger, P. Schmidt, and R. Smith, A surrogate model of gravitational waveforms from numerical relativity simulations of precessing binary black hole mergers, *Phys. Rev. D* **95**, 104023 (2017).
- [61] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, C. D. Ott, M. Boyle, L. E. Kidder, H. P. Pfeiffer, and B. Szilágyi, Numerical relativity waveform surrogate model for generically precessing binary black hole mergers, *Phys. Rev. D* **96**, 024058 (2017).
- [62] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, Surrogate models for precessing binary black hole simulations with unequal masses, *Phys. Rev. Res.* **1**, 033015 (2019).
- [63] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW190412: Observation of a binary-black-hole coalescence with asymmetric masses, *Phys. Rev. D* **102**, 043015 (2020).
- [64] M. Hannam *et al.*, General-relativistic precession in a black-hole binary, *Nature (London)* **610**, 652 (2022).
- [65] V. Varma, S. Biscoveanu, T. Islam, F. H. Shaik, C.-J. Haster, M. Isi, W. M. Farr, S. E. Field, and S. Vitale, Evidence of large recoil velocity from a black hole merger signal, *Phys. Rev. Lett.* **128**, 191102 (2022).
- [66] J. C. Bustillo, S. H. W. Leong, and K. Chandra, GW190412: Measuring a black-hole recoil direction through higher-order gravitational-wave modes, [arXiv:2211.03465](https://arxiv.org/abs/2211.03465).
- [67] J. C. Bustillo, A. del Rio, N. Sanchis-Gual, K. Chandra, and S. H. W. Leong, Testing mirror symmetry in the universe with LIGO-Virgo black-hole mergers, [arXiv:2402.09861](https://arxiv.org/abs/2402.09861).
- [68] T. Islam, A. Vajpeyi, F. H. Shaik, C.-J. Haster, V. Varma, S. E. Field, J. Lange, R. O’Shaughnessy, and R. Smith, Analysis of GWTC-3 with fully precessing numerical relativity surrogate models, [arXiv:2309.14473](https://arxiv.org/abs/2309.14473).
- [69] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, L. E. Kidder, and H. P. Pfeiffer, Surrogate model of hybridized numerical relativity binary black hole waveforms, *Phys. Rev. D* **99**, 064045 (2019).
- [70] J. Yoo, V. Varma, M. Giesler, M. A. Scheel, C.-J. Haster, H. P. Pfeiffer, L. E. Kidder, and M. Boyle, Targeted large mass ratio numerical relativity surrogate waveform model for GW190814, *Phys. Rev. D* **106**, 044001 (2022).
- [71] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), GW190814: Gravitational waves from the coalescence of a 23 solar mass black hole with a 2.6 solar mass compact object, *Astrophys. J. Lett.* **896**, L44 (2020).
- [72] D. Williams, I. S. Heng, J. Gair, J. A. Clark, and B. Khamesra, Precessing numerical relativity waveform surrogate model for binary black holes: A Gaussian process regression approach, *Phys. Rev. D* **101**, 063011 (2020).
- [73] T. Andrade, R. Gamba, and J. Trenado, Actively learning numerical relativity, [arXiv:2311.11311](https://arxiv.org/abs/2311.11311).
- [74] F. F. Freitas, C. A. R. Herdeiro, A. P. Morais, A. Onofre, R. Pasechnik, E. Radu, N. Sanchis-Gual, and R. Santos, Generating gravitational waveform libraries of exotic compact binaries with deep learning, [arXiv:2203.01267](https://arxiv.org/abs/2203.01267).
- [75] 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).
- [76] S. R. Brandt *et al.*, The Einstein toolkit, Zenodo, [10.5281/zenodo.5770803](https://zenodo.org/record/5770803) (2021).
- [77] T. Goodale, G. Allen, G. Lanfermann, J. Massó, T. Radke, E. Seidel, and J. Shalf, The Cactus framework and toolkit: Design and applications, in *Vector and Parallel Processing—VECPAR'2002, 5th International Conference, Lecture Notes in Computer Science* (Springer, Berlin, 2003).
- [78] J. D. Brown, P. Diener, O. Sarbach, E. Schnetter, and M. Tiglio, Turduckening black holes: An analytical and computational study, *Phys. Rev. D* **79**, 044023 (2009).
- [79] C. Reisswig, C. D. Ott, U. Sperhake, and E. Schnetter, Gravitational wave extraction in simulations of rotating stellar core collapse, *Phys. Rev. D* **83**, 064008 (2011).
- [80] H. Witek and M. Zilhão, CANUDA, <https://bitbucket.org/canuda/>.
- [81] M. Zilhão, H. Witek, and V. Cardoso, Nonlinear interactions between black holes and Proca fields, *Classical Quantum Gravity* **32**, 234003 (2015).
- [82] H. Witek, M. Zilhao, G. Bozzola, C.-H. Cheng, A. Dima, M. Elley, G. Ficarra, T. Ikeda, R. Luna, C. Richards, N. Sanchis-Gual, and H. Silva, Canuda: A public numerical relativity library to probe fundamental physics, Zenodo, [10.5281/zenodo.3565475](https://zenodo.org/record/3565475) (2023).
- [83] C. Cutler and E. E. Flanagan, Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral wave form?, *Phys. Rev. D* **49**, 2658 (1994).
- [84] T. A. Apostolatos, Search templates for gravitational waves from precessing, inspiraling binaries, *Phys. Rev. D* **52**, 605 (1995).
- [85] L. S. Finn, Detection, measurement and gravitational radiation, *Phys. Rev. D* **46**, 5236 (1992).
- [86] C. R. Galley, ROMPY (2020).
- [87] A. Paszke *et al.*, Pytorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems 32* (Curran Associates, Inc., Vancouver, Canada, 2019), pp. 8024–8035.
- [88] R. J. E. Smith, G. Ashton, A. Vajpeyi, and C. Talbot, Massively parallel Bayesian inference for transient gravitational-wave astronomy, *Mon. Not. R. Astron. Soc.* **498**, 4492 (2020).
- [89] J. S. Speagle, Dynesty: A dynamic nested sampling package for estimating bayesian posteriors and evidences, *Mon. Not. R. Astron. Soc.* **493**, 3132 (2020).
- [90] [10.54499/UIDB/04106/2020](https://zenodo.org/record/1054499).
- [91] [10.54499/UIDP/04106/2020](https://zenodo.org/record/1054499).
- [92] [10.54499/PTDC/FISAST/3041/2020](https://zenodo.org/record/1054499).
- [93] [10.54499/CERN/FIS-PAR/0024/2021](https://zenodo.org/record/1054499).
- [94] [10.54499/2022.04560.PTDC](https://zenodo.org/record/1054499).
- [95] R. Abbott *et al.*, Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo, *SoftwareX* **13**, 100658 (2021).
- [96] The LIGO Scientific Collaboration, The Virgo Collaboration, and The KAGRA Collaboration, Open data from the third observing run of LIGO, Virgo, KAGRA and GEO, *Astrophys. J. Suppl. Ser.* **267**, 29 (2023).