Rapid eccentric spin-aligned binary black hole waveform generation based on deep learning

Ruijun Shi,[1,2] Yue Zhou,[3] Tianyu Zhao,[4] Zhixiang Ren,[3, *] and Zhoujian Cao[1, 2, 5, †]

[1]*School of Physics and Astronomy, Beijing Normal University, Beijing, 100875, China*
[2]*Institute for Frontiers in Astronomy and Astrophysics,*
Beijing Normal University, Beijing, 102206, China
[3]*Peng Cheng Laboratory, Shenzhen 518055, China*
[4]*Center for Gravitational Wave Experiment, National Microgravity Laboratory,*
Institute of Mechanics, Chinese Academy of Sciences, Beijing, 100190, China
[5]*School of Fundamental Physics and Mathematical Sciences,*
Hangzhou Institute for Advanced Study, UCAS, Hangzhou, 310024, China
(Dated: November 25, 2024)

Accurate waveform templates of binary black holes (BBHs) with eccentric orbits are essential for the detection and precise parameter estimation of gravitational waves (GWs). While `SEOBNRE` produces accurate time-domain waveforms for eccentric BBH systems, its generation speed remains a critical bottleneck in analyzing such systems. Accelerating template generation is crucial to data analysis improvement and valuable information extraction from observational data. We present `SEOBNRE_AIq5e2`, an innovative AI-based surrogate model that crafted to accelerate waveform generation for eccentric, spin-aligned BBH systems. `SEOBNRE_AIq5e2` incorporates an advanced adaptive resampling technique during training, enabling the generation of eccentric BBH waveforms with mass ratios up to 5, eccentricities below 0.2, and spins $|\chi_z|$ up to 0.6. It achieves an impressive generation speed of 4.3 ms per waveform with a mean mismatch of 1.02×10^{-3}. With the exceptional accuracy and rapid performance, `SEOBNRE_AIq5e2` emerges as a promising waveform template for future analysis of eccentric gravitational wave data.

I. INTRODUCTION

The detection of the first gravitational wave (GW), GW150914 [1–3], has unveiled a new perspective on understanding the universe. So far, GWTC-3 has released nearly one hundred compact binary coalesce (CBC) events [4–6]. Traditional GW data analysis has primarily focused on waveforms that are generated by binary black hole (BBH) systems in circular orbits. In ground-based GW detection, we usually assume that gravitational waves have been circularized due to the effect of gravitational radiation. This assumption is consistent with the current GW detection results [4–6]. However, in recent years, there has been an increasing attention on detecting signs of eccentricity in present GW signals [7–14].

Currently, the analysis of GW190521 provides evidence of eccentricity residuals [15–18]. Eccentricity carries critical information about the astrophysical environments in which compact binaries form and evolve [19–23]. In certain extreme conditions, such as those found in dense stellar clusters [19], near supermassive black holes [24] and triple stars systems [25], binary black hole systems can retain significant orbital eccentricity as they enter the GW frequency bands. They are detectable by ground- and space-based GW detectors. Future space-based GW detection missions, including LISA [26], Taiji[27], and Tianqin[28], are expected to detect a broader range of GW signals. They have the potential to reveal a significant

population of BBHs with eccentric orbits [29, 30]. These observations will provide new insights into the formation channels and dynamic interactions that lead to eccentric compact binaries, deepening our understanding of the astrophysical processes in high-density environments [19].

The GW detection and parameter estimation from GW detectors rely on a high-accuracy waveform model [31]. Numerical relativity (NR) provides the most accurate waveform template, but NR requires a lot of computing resources and cannot be directly used in GW data analysis. The approximate templates used for data analysis are mainly effective-one-body (EOB) model [32] and IMRPhenom model [33, 34]. And the widely used eccentricity BBH waveform templates include `SEOBNRE` [35–37], `SEOBNRv4EHM` [38, 39], `TEOBResumS-Dalí` [40–42], `ENIGMA` [43, 44], `TaylorF2ECC` [45, 46], `EccentricFD`, and `EccentricTD` [47–50]. Compared to circular orbit waveforms, the evolution of waveforms in eccentric orbits is more complex. `SEOBNRE` is one of the most precise eccentric GW waveform models with a precision adequate for GW data analysis. At total mass $M_t = 60M_\odot$ and $f_0 = 20$Hz, the CPU generation speed of `SEOBNRE` is 2.1s, the generation speed becomes the main bottleneck when about one million waveforms are required to estimate a single eccentric BBH.

The application of graphics processing units (GPUs) holds considerable promise for substantially improving computational efficiency in the field of gravitational astronomy [51–53]. The IMRPhomon models such as BBHx [51, 54] can be readily implemented on GPU architectures. Nevertheless, the EOB models necessitate the resolution of a set of ordinary differential equations, mak-

* Corresponding author; renzhx@pcl.ac.cn
† Corresponding author; zjcao@bnu.edu.cn

ing their incorporation into parallel processing systems more complex. To address the CPU computational challenges, data-driven surrogate models have been developed to significantly accelerate this process. The surrogate models utilize machine learning methodologies, including the greedy algorithm [55] and proper orthogonal decomposition [56, 57], to develop a reduced order model (ROM). These surrogate models for NR and EOB waveforms provide computationally efficient waveform representations while maintaining a high level of accuracy [58–61]. SEOBNRE_S[61] and NRSur2dq1Ecc [60] constructed the reduced-order surrogate model for eccentric BBH waveform.

The rapid advancement of deep learning technologies in recent years has led to extensive applications [62] in GW detection [63–66], denoise[67, 68], parameter estimation [69–71], as well as in waveform generation [72–81]. The surrogate model based on deep learning enables rapid EOB waveform generation on GPU architectures. References [72–76] used an artificial neural network (ANN) to map GW parameters into their basis coefficients. However, these frameworks still rely on ROM techniques, which will be difficult to construct because they need a higher-order basis for the eccentricity waveform. Liao et al. [77] employed a conditional autoencoder to learn waveforms' latent space. However, this method needs mapping parameters to latent space, resulting in some loss of accuracy.

To accelerate the waveform generation for SEOBNRE, we propose a rapid generation framework SEOBNRE_AIq5e2 (Figure 1). In this framework, the waveforms are resampled to 1024 points during the training phase, and the model is trained on the resampled amplitude and phase. The resampled waveform is standardized to a fixed length of 1024 sampling points, which is unable to reconstruct to its original length directly. To address this issue, SEOBNRE_AIq5e2 uses a MLP to map waveform parameters to the waveform length, followed by an interpolation process to revert to the original waveform length. Unlike other AI surrogate models, we do not need to employ the ROM method to generate low-rank basis representations of the waveform, our model maps waveform parameters to waveforms directly. We utilize a single model to produce a shorter waveform that retains a substantial amount of the original waveform information. The model can generate waveforms rapidly with the reference frequency $Mf = 0.06$, mass ratio $q \in [1, 5]$, eccentricity $e \in [0.0.2]$, and spin $\chi_{1,2} \in [-0.6, 0.6]$, achieving an average mismatch of 1.02×10^{-3} and a signal waveform generation speed of 4.3 ms, which makes it possible for rapid parameter estimation and detection of the eccentricity BBHs.

The rest of this paper is organized as follows: Section II and III describe data generation and SEOBNRE_AIq5e2 architecture, respectively. In Section IV, we present the SEOBNRE_AIq5e2 waveform accuracy and speed results and discuss the future development directions. Finally, Section V highlights our findings based on the results.

II. DATASET CONSTRUCTION

A. Waveform generation

The training dataset for this study was based on the SEOBNRE waveform model. SEOBNRE is an advanced model in the family of EOB approaches designed to incorporate features of NR simulations for eccentric, spinning binary systems with 2PN orbital radiation-reaction forces. It is particularly well-suited for accurately modeling the inspiral, merger, and ringdown phases of BBH coalescences, including systems with non-negligible eccentricity and spins [35–37].

Each waveform was generated in the form of the two polarizations, h_+ and $h_\times$, over a time interval that captures the full inspiral, merger, and ringdown. To simplify model training, we separate the learning of the waveform's amplitude and phase components:

$$h(t, \lambda) = h_+(t, \lambda) - ih_\times(t, \lambda)$$
$$= \mathcal{A}(t, \lambda)e^{-i\phi(t,\lambda)}, \quad (1)$$

where $\mathcal{A}(t, \lambda)$ represents the waveform amplitude, $\phi(t, \lambda)$ is the waveform phase, and λ represents the GW source parameter. By decomposing the waveform in this way, each component is processed independently, allowing the model to focus on specific features of the amplitude and phase without interference from one another. The parameter space of the train and test dataset is shown in Table I. In this study, we restrict our analysis to the dominant $(2, 2)$ mode of the GW signal.

During dataset generation, the dimensionless reference frequency was fixed at $Mf = 0.06$, sample rate is set to 4096Hz, and the total mass of the BBH was set to $60M_\odot$ ($f_0 \approx 20$Hz). The value of Mf is fixed, and it can be rescaled across different total masses. This guarantees that the generated waveforms span the frequency bands of the ground-based GW detectors and the temporal length appropriate for the detection and analysis of stellar.

B. Merger time alignment

Previous approaches constructed a data-driven surrogate model by aligning merging times. The eccentric GW waveform has a far more complex waveform than the non-ecc waveform, and waveforms with varied parameters exhibit apparent fluctuations. Figure 2(a) shows fluctuations in the parameter space $e_0 - t$. Unfortunately, data-driven algorithms prefer to learn smooth components, and these fluctuation components are more difficult to fit. For example, the SEOBNRE_S model employs over 300 basic functions to reach an accuracy level comparable to that achieved with only 100 bases in non-eccentric waveform models NRSur7dq4 at 7 dimension parameters [59].

At the same time, given a fixed-length output structure that uses a non-recurrent neural network architecture,
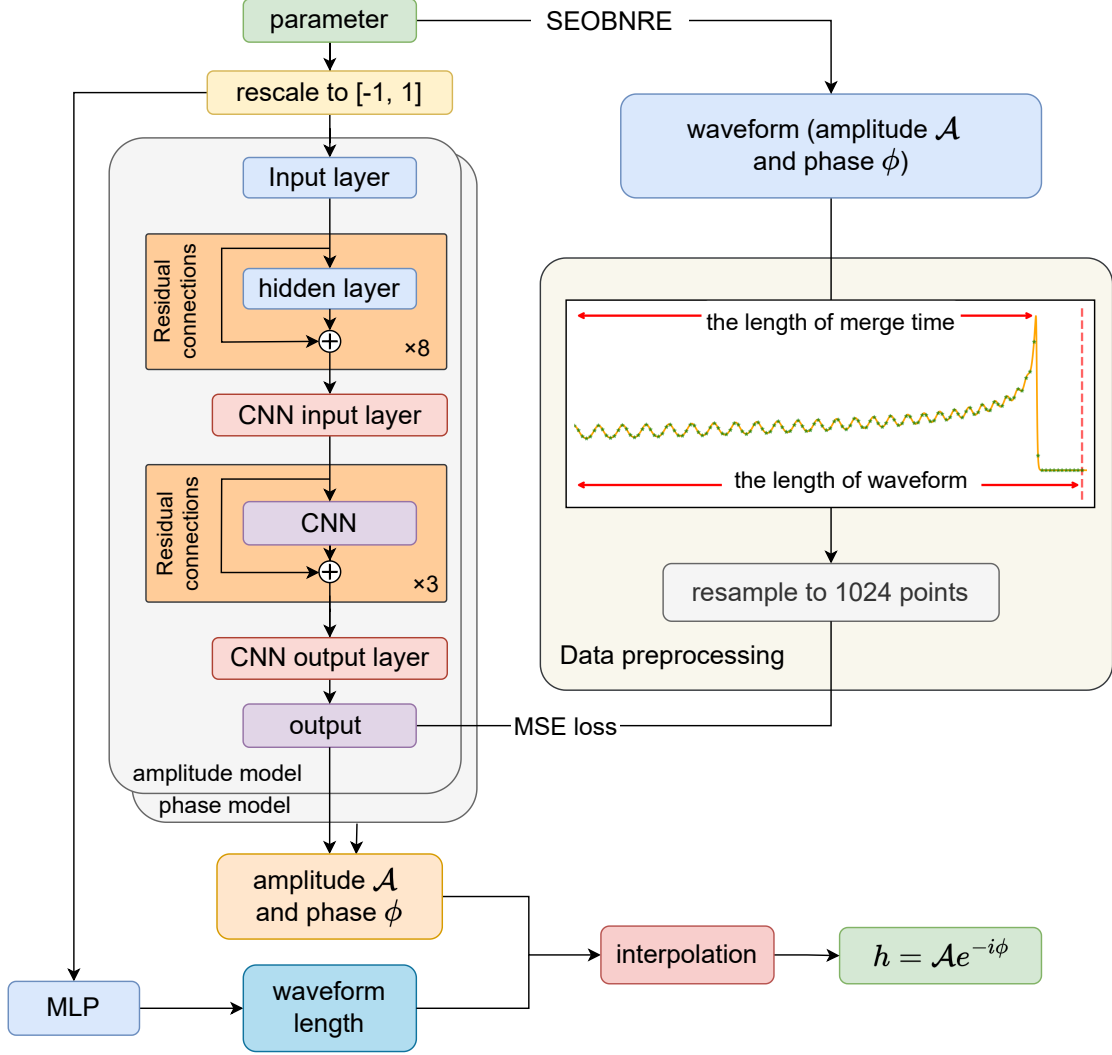
Figure 1. **The overview of data pre-processing and model architecture.** The data pre-processing section (right) illustrates the steps taken to prepare training samples, including waveform parameter normalization and length resampling.

Table I. Parameter distribution of training dataset and test dataset

Parameter	Description	Parameter distribution
M_{tot}	Total mass of BBHs $m_1 + m_2$	Fix at $60M_\odot$
q	Mass ratio $\frac{m_1}{m_2}$ and $q \geq 1$	Uniform $[1, 5]$
e_0	eccentricity at f_0	Uniform $[0, 0.2]$
χ_1, χ_2	Spin parameters of two black holes	Uniform $[-0.6, 0.6]$

the shorter waveforms of the model require zero-padding to match the output length. This approach introduces an imbalance in the training sample distribution due to inconsistencies in waveform length across different parameters. If the waveform is cut to the smallest length in the training set, the gravitational wave information is lost. Furthermore, neither the zero-padding nor waveform

truncation can alleviate the fluctuation phenomena in parameter space.

We employ an adaptive resampling technique to address sample imbalance and reduce information loss from truncation. This approach resamples waveforms of different lengths to a consistent number of sampling points (Figure 2(b)), ensuring consistency across samples while
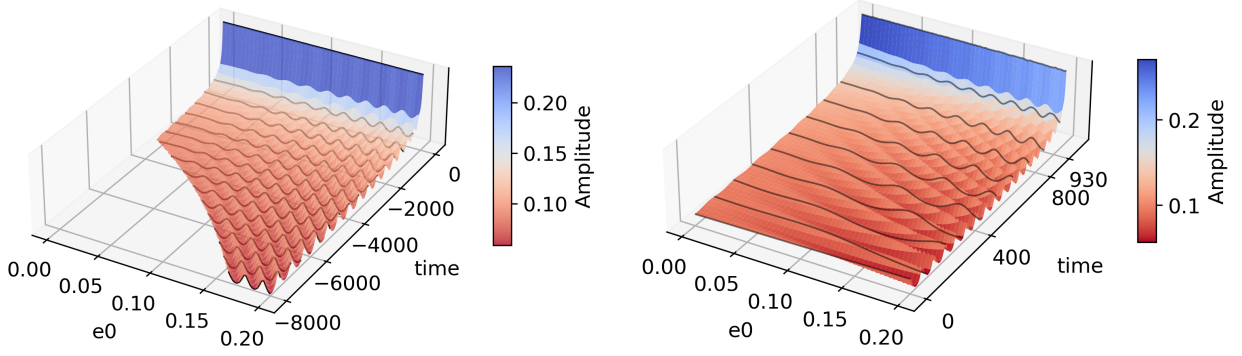
(a) Merger time alignment, merge time is at zero point.   (b) Waveform resampling, initial evolution time is at zero point.

Figure 2.   **The surface of $\mathcal{A}(e_0; t)$ for the (2,2) mode.** Here we set $q = 2.2$ and $\chi_{1,2} = 0$. The black lines that cover the surface indicate the fluctuation of the amplitude $\mathcal{A}_t(e_0)$ in the parameter space at a fixed time point, and the time axis is the corresponding sampling point. (a) shows fluctuations in the parameter space e_0. (b). waveforms are resampled to same length to ensure uniformity and crucial information preservation.

preserving essential information with minimal loss. Moreover, the fluctuation in parameter space is alleviated in the early sample points. Here we use cubic spline interpolation for resampling. In section II C we will introduce how to resample the waveform.

C.   Waveform resample and interpolation

To standardize the length of waveforms across different parameters, each waveform was rescaled to a length corresponding to 1.1 times from the t_0 to the merger time t_m, which is called the original waveform duration t_{orig}:

$$t_{\text{orig}} = 1.1 \times (t_m - t_0). \tag{2}$$

Because the sample rate is constant at 4096 Hz, for simplicity, t_{orig} is an integer, referring to the number of sampling points.

Subsequently, each waveform is resampled to 1024 points using cubic spline interpolation, which ensures consistent data representation across the dataset and preserves the key features of the signal while reducing the computational complexity. The resampling points are:

$$t_{\text{resample}} = \frac{i}{N-1} \cdot (t_{\text{orig}} - 1), \quad i = 0, 1, \dots, 1023, \tag{3}$$

where N refers to the length after resample, and we set $N = 1024$.

The model will generate a sequence of 1024 points, which must be returned to the original time series. To recover the original length, the model will output the expected $\hat{t}_{\text{orig}}$ and use cubic spline interpolation to the

recovered time series t_{interp}:

$$t_{\text{interp}} = \frac{i}{\hat{t}_{\text{orig}} - 1} \cdot 1023, \quad i = 0, 1, \dots, \hat{t}_{\text{orig}} - 1. \tag{4}$$

This approach also encounters high-frequency fluctuations within the parameter space before the merger time; however, in Figure 2(b) these fluctuations are significantly mitigated during the early stages of evolution. This reduction enables the model to efficiently learn the overall structure of the waveform and capture the initial signal characteristics. With further training to incorporate higher-frequency details, the model achieves an output of high-accuracy waveforms. Figure 1 illustrates the preprocessed waveforms. This resampling approach guarantees efficient training of the AI model by maintaining a uniform input size.

Our model produces two time series outputs: amplitude and phase, along with the corresponding waveform duration. The model output time series are interpolated to the defined waveform duration, ensuring consistency with the physical duration of the waveform. The duration of the final output waveform varies depending on the input waveform parameters.

The parameters of the training set and the test set are randomly sampled in the parameter space. A total of 500,000 waveforms were generated for the training set, while the additional 10,000 waveforms were generated for the test set.

III.   SEOBNRE_AIQ5E2 MODEL

The `SEOBNRE_AIq5e2` model inputs the GW source parameter and outputs the generated waveform that is

derived from three characteristics: the amplitude, phase, and length of the waveform. This structured approach enables a comprehensive representation of the GW, facilitating accurate modeling of its characteristics based on the input source parameters.

A. Amplitude and phase generation model

The `SEOBNRE_AIq5e2` model combines a multi-layer perception (MLP) with a convolutional neural network (CNN). This hybrid approach leverages both the MLP's capability to learn complex parameter-space mappings and the CNN's efficiency in modeling temporal correlations. Respective models are trained for amplitude $\mathcal{A}(t, \lambda)$ and phase $\phi(t, \lambda)$ prediction.

Normalizing the input parameters.— Since the value ranges between different parameters are inconsistent, a linear mapping is applied to each parameter to standardize the input. Specifically, each parameter p is linearly mapped to the range $[-1, 1]$ using the following transformation:

$$p_{\mathrm{norm}} = 2 \times \frac{p - p_{\mathrm{min}}}{p_{\mathrm{max}} - p_{\mathrm{min}}} - 1, \tag{5}$$

where p_{min} and p_{max} represent the upper and lower limits of the parameter range of parameter p respectively. This normalization ensures that all input parameters are on a consistent scale, enabling a more stable and efficient training of the model.

MLP.— The MLP layer used in this model is designed to map input parameters to high-dimensional latent representations. It facilitates the model to capture complex relationships between input parameters and waveform outputs. The MLP consists of an input layer, followed by 8 hidden layers, each with residual connections and ReLU activation.

The input to the MLP is a 4-dimensional vector that represents the parameters of the wave source. The input layer maps this 4-dimensional input into a 1024-dimensional hidden space. The transformation applied in the input layer can be represented as:

$$\mathbf{h}^{(0)} = \mathrm{ReLU}\left(\mathbf{W}^{(0)}\mathbf{p} + \mathbf{b}^{(0)}\right), \tag{6}$$

where $\mathbf{p} \in \mathbb{R}^4$ is the input vector, $\mathbf{W}^{(0)} \in \mathbb{R}^{1024 \times 4}$ and $\mathbf{b}^{(0)} \in \mathbb{R}^{1024}$ are the weight matrix and bias vector of the input layer, respectively. Following the input layer, each hidden layer has a hidden size of 1024. Residual connections are incorporated into each hidden layer to stabilize gradient flow and model performance. For each hidden layer l, where $l = 1, \ldots, 8$, the output is computed as follows:

$$\mathbf{h}^{(l)} = \mathrm{ReLU}\left(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}\right) + \mathbf{h}^{(l-1)}, \tag{7}$$

where $\mathbf{W}^{(l)} \in \mathbb{R}^{1024 \times 1024}$, $\mathbf{b}^{(l)} \in \mathbb{R}^{1024}$ are the bias vector for the l-th hidden layer, $\mathbf{h}^{(l-1)} \in \mathbb{R}^{1024}$ is the output of

the previous layer, and the residual connection adds this to the transformed output of the current layer.

CNN.— The output of the MLP is subsequently inserted into the CNN. Firstly, the MLP output is transformed into a 32-channel representation. This multichannel expansion enhances the model's capacity to capture intricate waveform features. Subsequently, three consecutive convolutional layers with residual connection are employed to smooth the waveform across channels, refining the feature representation and preserving essential temporal patterns. Finally, an additional convolutional layer condenses the 32-channel output back into a single channel, yielding a processed waveform that retains both the structural integrity and details for high-accuracy waveform reconstruction. This hierarchical approach allows the CNN to learn intricate temporal dynamics associated with the gravitational waveforms, enhancing the model's generation accuracy.

Additionally, each convolutional layer utilizes the ReLU activation function and instance normalization (InstanceNorm) to promote stable training and enhance feature extraction across the network. The InstanceNorm [82] is computed as:

$$\hat{\mathbf{h}}^{(l)} = \frac{\mathbf{h}^{(l)} - \mu}{\sigma} + \beta, \tag{8}$$

where μ and σ are the mean and standard deviation of the features for each instance, respectively. This normalization helps to reduce covariate shift during training, leading to faster convergence and improved model stability.

B. Waveform-length model

To address the challenge of varying waveform lengths across different parameter configurations, we resample all waveforms at the same length. However, this standardization renders the actual waveform length unknown for each case. To predict the original waveform length, we employ a 4-layer MLP model with a hidden layer size of 1024. This MLP model is trained separately to estimate the waveform length based on the waveform parameter, enabling the framework to subsequently apply interpolation. Specifically, we use the cubic spline interpolation, a method known for its smoothness and efficiency in approximating continuous functions, to recover the waveform. This approach ensures consistency in input dimensions while retaining essential information about the waveform's natural length.

C. Training and Verification

During training, the Adam optimizer [83] is used to optimize the model. We use signal RTX 4090 GPU to train the amplitude, phase and waveform length model in this study for approximately 12 hours, respectively. The learning rate was initialized at 10^{-4} and decayed by a

factor of 0.95 every 50 epochs to gradually reduce the step size and enhance convergence stability over the course of training.

To assess the MLP-CNN model's capacity to generate GW waveform, we measure its recovery ability by comparing `SEOBNRE_AIq5e2` waveform with SEOBNRE waveform. Eq. 9 calculates the overlap $\mathcal{O}$ between the target waveform and the generated waveform using the `pycbc` library [84]. The overlap value ranges from 0 to 1, with higher values indicating greater similarity between the generated waveform and the target waveform.

$$\mathcal{O}(h_1, h_2) = \max_{t_c, \phi_c} \left(\hat{h}_1 | \hat{h}_2 \right)^{1/2}, \tag{9}$$

$$\hat{h} = \frac{h}{\sqrt{(h|h)}}, \tag{10}$$

where $(a|b)$ represents the inner product:

$$(h_1|h_2) = 4\mathrm{Re} \int_{f_{\min}}^{f_{\max}} \frac{\tilde{h}_1^*(f)\tilde{h}_2(f)}{S_n(f)} df, \tag{11}$$

where S_n is one-sided power spectral density (PSD) of the detector noise and is set to 1. Finally, the mismatch between two waveform is defined as

$$\mathcal{M}(h_1, h_2) = 1 - \mathcal{O}(h_1, h_2) \tag{12}$$

IV. RESULTS

In this section, we evaluate the model's performance based on two primary metrics: waveform mismatch and generation speed. Mismatch is assessed to quantify the accuracy of `SEOBNRE_AIq5e2` waveforms relative to the target waveforms, providing insights into accuracy across a range of parameter configurations. Generation speed is examined to demonstrate the efficiency of the `SEOBNRE_AIq5e2` waveform model.

A. Mismatch Reuslts

Figure 3 shows the mismatch between the targets and generated waveforms of `SEOBNRE_AIq5e2` at different total masses. In the test set ($M_t = 60M_\odot$), the mean mismatch is 1.02×10^{-3}, with mismatch values ranging from a minimum of 3.63×10^{-5} to a maximum of 3.31×10^{-2}. Additionally, Figure 4 illustrates showcases of waveform generation by our AI model, demonstrating its strong capability to replicate target waveforms across wide GW parameter range accurately.

Using a reference total mass of $M_t = 60M_\odot$, we evaluate the mismatch for various total masses, specifically $M_t = (30, 60, 120, 180, 240)M_\odot$ and the violin plots are shown in Figure 3. The mean mismatch remains relatively stable and low across different mass scalings, with a slight
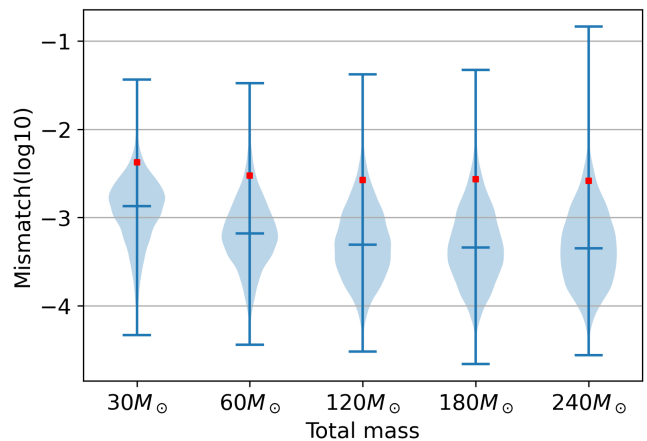

Figure 3. **Distribution of mismatches between `SEOBNRE` and `SEOBNRE_AIq5e2` on the test set.** During test we rescale the waveforms of $M_t = 60M_\odot$ to different total masses. The red mark represents the 95% percentile of the mismatch.

improvement in mismatch observed at higher total masses compared to lower total masses. We observe that the maximum mismatch value increases when total masses become large. This tendency occurs as higher-mass systems are more sensitive to the merger stage [73], where any tiny errors in generation waveforms can be amplified.

Figure 5 displays the distribution in parameter space of waveforms with mismatch values exceeding the 95th percentile at $M_t = 60M_\odot$. The histogram of Figure 5 shows that the mismatch is larger where the mass ratio is higher or lower. Simultaneously, the higher the eccentricity e_0 and spin $\chi_{1,2}$, the bigger the mismatch. While the scatter plot is generally uniformly distributed, there are some tendencies for clustering in certain regions. In the mismatch scatter plot across the q-e_0 parameter space, regions of higher mismatch are predominantly located at larger eccentricities, larger or smaller mass ratios. Similarly, there is a cluster in high-spin areas. A similar pattern of distribution is observed even when the waveforms are rescaled to different total masses. As the initial eccentricity and mass ratio increase, the waveform duration extends, adding complexity to model training and requiring a greater interpolation factor, both of which contribute to higher mismatch values.

B. Waveform generation speed

Compared to EOB waveform models, a significant advantage of data-driven waveforms lies in their substantially faster generation speed. On a single RTX 4090 GPU, the model achieves a waveform generation speed of 4.3 ms (Table II). `SEOBNRE_AIq5e2` is also competitive in terms of frequency domain waveform generation speed and can provide a more comprehensive eccentric waveform. There is nearly $500\times$ faster than `SEOBNRE` waveform generation at one CPU core (Intel 2.3GHz Xeon 8336C). The model
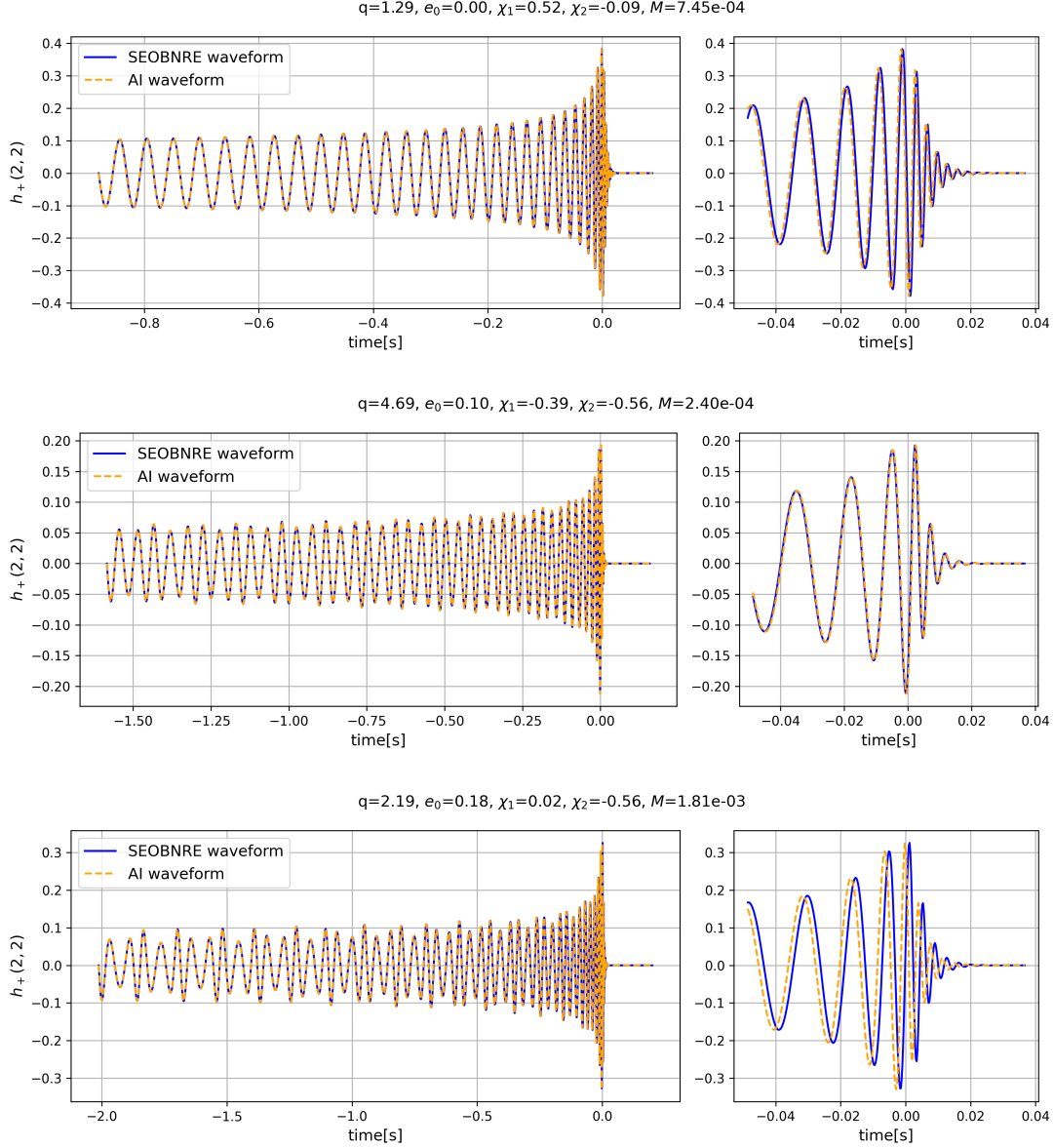
Figure 4. **Comparison of generated waveforms between `SEOBNRE_AIq5e2` and `SEOBNRE`.** The figure demonstrates the waveform generation capability of `SEOBNRE_AIq5e2` model across a wide range of parameter configurations, where M stands for mismatch value.

demonstrates efficient training speeds. With a training set of 50,000 samples, the time required to train a single model for amplitude, phase, or waveform length is approximately 12 hours over 1,500 epochs.

Another advantage of using GPUs is their high degree of parallelism. In Table III, we compared the generation speed of the single GPU and the 64-core CPUs ($2\times$Intel 2.3GHz Xeon 8336C) with batch=(1, 10, 100, 1000). To benchmark generation speed against `SEOBNRE`, we employ multiple processes for waveform generation, with the batch size in Table III representing the number of concurrent processes used. Even at larger batch sizes, the model has remarkable acceleration performance, which

is a great benefit for Bayesian inference. For example, `SEOBNRE_AIq5e2` generates the required 1000 waveforms in 0.812 seconds using parallel-tempered MCMC tools [85], whereas 64-core CPUs take around 39.34 s. This means `SEOBNRE_AIq5e2` is about $48\times$ faster when run in highly parallel setups. The AI model generates waveforms at a consistent speed across different parameters and efficiently handles longer waveforms in extreme cases. Although the interpolation process becomes slower with increasing waveform length, the time required for interpolation constitutes only 1/10 to 1/100 of the total time that is needed to generate the AI-produced part. For the case of generating a large number of waveforms in
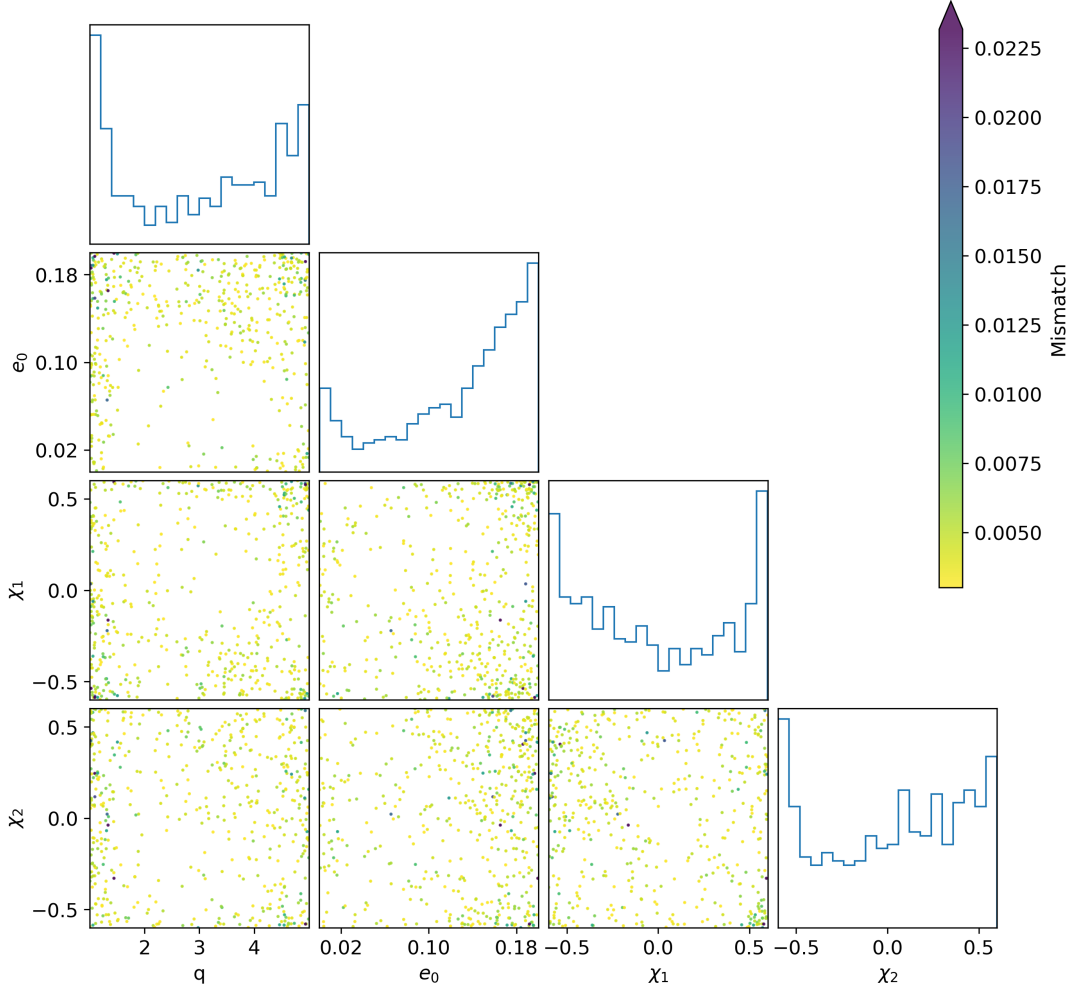
Mismatch ≥ 0.002998

Figure 5. **Distribution of mismatches between target waveforms and AI-generated waveforms on the test set**.
Here is a scatter plot of mismatch above the 95% quantile at $M_t = 60 M_\odot$ in the (q, e_0, χ_1, χ_2) parameter space. Higher mass
ratio, eccentricity, and spin all exhibit higher mismatch values. In addition, the area with a low mass ratio will have a larger
mismatch value. The other total masses are similar to the $M_t = 60 M_\odot$.

parallel, the efficiency of performing parallel interpolation
calculations can be further improved on this basis (Table
III).

V. CONCLUSION AND DISCUSSION

In this study, we present the `SEOBNRE_AIq5e2` model,
establishing a novel framework for waveform surrogate
modeling based on deep learning. This is the first model to
use deep learning in the field of eccentric waveform gener-
ation. Through this approach, `SEOBNRE_AIq5e2` advances
the efficiency and accuracy of eccentric waveform genera-
tion, offering a powerful tool for enhancing gravitational
wave data analysis. At the same time, `SEOBNRE_AIq5e2`

Table II. **Time that required to generate a single wave-
form.** Here, we choose the average generation speed of 100
waveforms. The waveform production settings are the same
as in Section II. For the frequency domain waveform, we set
$f_{min} = 20 Hz$ and $\Delta f = 1/8 Hz$.

waveform	speed
*`SEOBNRE_AIq5e2`(GPU)	4.288 ms
*`SEOBNRE_AIq5e2`(CPU)	9.856 ms
*`SEOBNRE`	2.133 s
*`EccentricTD`	61.948 ms
**`TaylorF2Ecc`	0.238 ms
**`EccentricFD`	2.465 ms

* Time domain waveform
** Frequency domain waveform

Table III. **The computational efficiency of `SEOBNRE_AIq5e2` under different batch sizes.** We evaluated the generation speed of `SEOBNRE_AIq5e2` on both GPU and CPUs, and `SEOBNRE` generation speed across varying process counts on the 64-core CPUs.

Batch size	GPU	CPU	SEOBNRE	speed up (SEOBNRE/GPU)
1	4.29e-3 s	9.86e-3 s	2.133 s	497
10	0.0123 s	0.0277 s	2.824 s	229
100	0.0819 s	0.2170 s	4.514 s	55
1000	0.8123 s	1.2404 s	39.340 s	48

is readily portable to the CPU, and the generation speed of a 64-core CPU is merely twice as a GPU. This implies that users without a GPU may assure their generating efficiency under the CPU.

In contrast to existing surrogate methods, this paper introduces a resampling-interpolation approach to enable variable-length waveform generation within a data-driven framework. This approach effectively addresses the frequency principle limitation [86] in deep learning, which significantly reduces oscillations in the parameter space. `SEOBNRE_AIq5e2` achieves a mean mismatch of 1.02×10^{-3} under conditions of $Mf = 0.06$, $1 < q < 5$, $0 < e < 0.2$, and spin parameters $-0.6 < \chi_{1,2} < 0.6$, demonstrating its robustness and accuracy across diverse waveform conditions. The model demonstrates robust accuracy across varying total mass, and we have shown that cubic spline interpolation effectively maintains the precision of waveform generation.

Compared to the `SEOBNRE` waveform, `SEOBNRE_AIq5e2` model achieves significantly faster generation speeds, requiring only 4.3 ms to produce a single waveform, with an improvement of two orders of magnitude over SEOBNRE. Leveraging GPU-based high-performance computing, the model enables simultaneous generation of multiple waveforms, reducing computational costs substantially and enhancing efficiency in large-scale waveform synthesis. Methods such as importance sampling, density estimation, and PTMCMC require the parallel generation of numerous waveforms for each evaluation, highlighting the advantages of GPU high-performance computing. The `SEOBNRE_AIq5e2` model is well-suited to these demands, as it significantly reduces waveform generation time. This model has the potential to greatly shorten parameter estimation times and enhance the efficiency of parameter estimation for eccentricity waveforms.

In future work, we aim to broaden the parameter space and incorporate higher modes within our model to enhance its applicability for the detection and parameter estimation of BBHs in eccentric orbits. Additionally, we will explore the potential of fine-tuning the `SEOBNRE_AIq5e2` model using NR waveforms to further improve accuracy. We will also look at using AI-driven interpolation approaches to enable a completely differentiable waveform template, allowing us to leverage gradient-based MCMC techniques like Hamiltonian Monte Carlo [87] to estimate parameters more efficiently.

ACKNOWLEDGMENTS

[1] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al., Physical Review Letters **116**, 061102 (2016).

[2] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al., Physical Review X **6**, 041015 (2016).

[3] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al., Physical Review D **93**, 122003 (2016).

[4] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, et al., Physical Review X **9**, 031040 (2019).

[5] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, et al., Physical Review X **11**, 021053 (2021).

[6] R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, N. Adhikari, R. X. Adhikari, V. B. Adya, C. Affeldt, D. Agarwal, et al., Physical Review X **13**, 041039 (2023).

[7] M. E. Lower, E. Thrane, P. D. Lasky, and R. Smith, Physical Review D **98**, 083028 (2018).

[8] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, et al., The Astrophysical Journal **883**,

149 (2019).

[9] I. M. Romero-Shaw, P. D. Lasky, and E. Thrane, Monthly Notices of the Royal Astronomical Society **490**, 5210 (2019).

[10] A. H. Nitz, A. Lenon, and D. A. Brown, The Astrophysical Journal **890**, 1 (2020).

[11] I. Romero-Shaw, P. D. Lasky, and E. Thrane, The Astrophysical Journal Letters **921**, L31 (2021).

[12] M. Favata, C. Kim, K. G. Arun, J. Kim, and H. W. Lee, Physical Review D **105**, 023003 (2022).

[13] A. Bonino, R. Gamba, P. Schmidt, A. Nagar, G. Pratten, M. Breschi, P. Rettegno, and S. Bernuzzi, Physical Review D **107**, 064024 (2023).

[14] Y. Xu and E. Hamilton, Physical Review D **107**, 103049 (2023).

[15] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, et al., Physical Review Letters **125**, 101102 (2020).

[16] I. Romero-Shaw, P. D. Lasky, E. Thrane, and J. C. Bustillo, The Astrophysical Journal Letters **903**, L5 (2020).

[17] V. Gayathri, J. Healy, J. Lange, B. O'Brien, M. Szczepańczyk, I. Bartos, M. Campanelli, S. Klimenko, C. O. Lousto, and R. O'Shaughnessy, Nature Astronomy **6**, 344 (2022).

[18] H. L. Iglesias, J. Lange, I. Bartos, S. Bhaumik, R. Gamba, V. Gayathri, A. Jan, R. Nowicki, R. O'Shaughnessy, D. M. Shoemaker, et al., The Astrophysical Journal **972**, 65 (2024).

[19] M. Zevin, J. Samsing, C. Rodriguez, C.-J. Haster, and E. Ramirez-Ruiz, The Astrophysical Journal **871**, 91 (2019).

[20] V. Cardoso, C. F. B. Macedo, and R. Vicente, Physical Review D **103**, 023015 (2021).

[21] M. Zevin, I. M. Romero-Shaw, K. Kremer, E. Thrane, and P. D. Lasky, The Astrophysical Journal Letters **921**, L43 (2021), arXiv:2106.09042 [astro-ph, physics:gr-qc].

[22] A. M. Knee, I. M. Romero-Shaw, P. D. Lasky, J. McIver, and E. Thrane, The Astrophysical Journal **936**, 172 (2022).

[23] Divyajyoti, S. Kumar, S. Tibrewal, I. M. Romero-Shaw, and C. K. Mishra, Physical Review D **109**, 043037 (2024).

[24] H. Tagawa, B. Kocsis, Z. Haiman, I. Bartos, K. Omukai, and J. Samsing, The Astrophysical Journal Letters **907**, L20 (2021).

[25] F. Antonini, S. Toonen, and A. S. Hamers, The Astrophysical Journal **841**, 77 (2017).

[26] P. Amaro-Seoane, H. Audley, S. Babak, J. Baker, E. Barausse, P. Bender, E. Berti, P. Binetruy, M. Born, D. Bortoluzzi, et al., "Laser Interferometer Space Antenna," (2017), arxiv:1702.00786 [astro-ph].

[27] W.-R. Hu and Y.-L. Wu, National Science Review **4**, 685 (2017).

[28] J. Luo, L.-S. Chen, H.-Z. Duan, Y.-G. Gong, S. Hu, J. Ji, Q. Liu, J. Mei, V. Milyukov, M. Sazhin, et al., Classical and Quantum Gravity **33**, 035010 (2016), arxiv:1512.02076 [astro-ph, physics:gr-qc].

[29] K. Breivik, C. L. Rodriguez, S. L. Larson, V. Kalogera, and F. A. Rasio, The Astrophysical Journal Letters **830**, L18 (2016).

[30] A. Nishizawa, A. Sesana, E. Berti, and A. Klein, Monthly Notices of the Royal Astronomical Society **465**, 4375 (2017).

[31] T. L. S. Collaboration, the Virgo Collaboration, B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya, et al., Classical and Quantum Gravity **37**, 055002 (2020), arXiv:1908.11170.

[32] A. Taracchini, A. Buonanno, Y. Pan, T. Hinderer, M. Boyle, D. A. Hemberger, L. E. Kidder, G. Lovelace, A. H. Mroué, H. P. Pfeiffer, et al., Physical Review D **89**, 061502 (2014).

[33] S. Husa, S. Khan, M. Hannam, M. Pürrer, F. Ohme, X. J. Forteza, and A. Bohé, Physical Review D **93**, 044006 (2016).

[34] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. J. Forteza, and A. Bohé, Physical Review D **93**, 044007 (2016).

[35] Z. Cao and W.-B. Han, Physical Review D **96**, 044028 (2017).

[36] X. Liu, Z. Cao, and Z.-H. Zhu, Classical and Quantum Gravity **39**, 035009 (2022).

[37] X. Liu, Z. Cao, and Z.-H. Zhu, Classical and Quantum Gravity **41**, 195019 (2024).

[38] M. Khalil, A. Buonanno, J. Steinhoff, and J. Vines, Physical Review D **104**, 024046 (2021).

[39] A. Ramos-Buades, A. Buonanno, M. Khalil, and S. Ossokine, Physical Review D **105**, 044035 (2022).

[40] D. Chiaramello and A. Nagar, Physical Review D **101**, 101501 (2020).

[41] A. Albertini, R. Gamba, A. Nagar, and S. Bernuzzi, Physical Review D **109**, 044022 (2024).

[42] A. Nagar, R. Gamba, P. Rettegno, V. Fantini, and S. Bernuzzi, Physical Review D **110**, 084001 (2024).

[43] E. A. Huerta, C. J. Moore, P. Kumar, D. George, A. J. K. Chua, R. Haas, E. Wessel, D. Johnson, D. Glennon, A. Rebei, et al., Physical Review D **97**, 024031 (2018).

[44] Z. Chen, E. A. Huerta, J. Adamo, R. Haas, E. O'Shea, P. Kumar, and C. Moore, Physical Review D **103**, 084018 (2021).

[45] B. Moore, M. Favata, K. G. Arun, and C. K. Mishra, Physical Review D **93**, 124061 (2016).

[46] B. Moore and N. Yunes, Classical and Quantum Gravity **36**, 185003 (2019).

[47] S. Tanay, M. Haney, and A. Gopakumar, Physical Review D **93**, 064031 (2016).

[48] V. Tiwari, S. Klimenko, N. Christensen, E. A. Huerta, S. R. P. Mohapatra, A. Gopakumar, M. Haney, P. Ajith, S. T. McWilliams, G. Vedovato, et al., Physical Review D **93**, 043007 (2016).

[49] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration, "LVK Algorithm Library - LALSuite," Free software (GPL) (2018).

[50] K. Wette, SoftwareX **12**, 100634 (2020).

[51] M. L. Katz, S. Marsat, A. J. K. Chua, S. Babak, and S. L. Larson, Physical Review D **102**, 023033 (2020).

[52] M. L. Katz, J.-B. Bayle, A. J. K. Chua, and M. Vallisneri, Physical Review D **106**, 103001 (2022), arxiv:2204.06633 [astro-ph, physics:gr-qc].

[53] S. H. Strub, L. Ferraioli, C. Schmelzbach, S. C. Stähler, and D. Giardini, Physical Review D **108**, 103018 (2023).

[54] M. L. Katz, Physical Review D **105**, 044055 (2022).

[55] S. E. Field, C. R. Galley, J. S. Hesthaven, J. Kaye, and M. Tiglio, Physical Review X **4**, 031006 (2014).

[56] A. Chatterjee, Current Science **78**, 808 (2000), 24103957.

[57] M. Tiglio and A. Villanueva, Living Reviews in Relativity **25**, 2 (2022).

[58] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, D. A. Hemberger, P. Schmidt, and R. Smith, Physical Review D **95**, 104023 (2017).

[59] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, Physical Review Research **1**, 033015 (2019).

[60] T. Islam, V. Varma, J. Lodman, S. E. Field, G. Khanna, M. A. Scheel, H. P. Pfeiffer, D. Gerosa, and L. E. Kidder, Physical Review D **103**, 064022 (2021).

[61] Q. Yun, W.-B. Han, X. Zhong, and C. A. Benavides-Gallego, Physical Review D **103**, 124053 (2021).

[62] T. Zhao, R. Shi, Y. Zhou, Z. Cao, and Z. Ren, "Dawning of a New Era in Gravitational Wave Data Analysis: Unveiling Cosmic Mysteries via Artificial Intelligence – A Systematic Review," (2023), arxiv:2311.15585 [astro-ph, physics:gr-qc].

[63] D. George and E. Huerta, Physics Letters B **778**, 64 (2018).

[64] D. George and E. A. Huerta, Physical Review D **97**, 044039 (2018).

[65] H. Gabbard, M. Williams, F. Hayes, and C. Messenger, Physical Review Letters **120**, 141103 (2018).

[66] T. Zhao, Y. Zhou, R. Shi, Z. Cao, and Z. Ren, Physical Review D **109**, 084054 (2024).

[67] H. Wang, Y. Zhou, Z. Cao, Z. Guo, and Z. Ren, Machine Learning: Science and Technology **5**, 015046 (2024).

[68] T. Zhao, R. Lyu, H. Wang, Z. Cao, and Z. Ren, Communications Physics **6**, 212 (2023).

[69] S. R. Green and J. Gair, Machine Learning: Science and Technology **2**, 03LT01 (2021).

[70] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, Physical Review Letters **127**, 241103 (2021).

[71] H. Wang, Z. Cao, Y. Zhou, Z.-K. Guo, and Z. Ren, Big Data Mining and Analytics **5**, 53 (2022).

[72] A. J. K. Chua, C. R. Galley, and M. Vallisneri, Physical Review Letters **122**, 211101 (2019), arXiv:1811.05491 [astro-ph, physics:gr-qc, stat].

[73] S. Khan and R. Green, Physical Review D **103**, 064015 (2021), arXiv:2008.12932 [gr-qc].

[74] P. Nousi, S.-C. Fragkouli, N. Passalis, P. Iosif, T. Apostolatos, G. Pappas, N. Stergioulas, and A. Tefas, Neurocomputing **491**, 67 (2022).

[75] L. M. Thomas, G. Pratten, and P. Schmidt, Physical Review D **106**, 104029 (2022).

[76] S.-C. Fragkouli, P. Nousi, N. Passalis, P. Iosif, N. Stergioulas, and A. Tefas, Applied Soft Computing **147**, 110746 (2023).

[77] C.-H. Liao and F.-L. Lin, Physical Review D **103**, 124051 (2021).

[78] J. Lee, S. H. Oh, K. Kim, G. Cho, J. J. Oh, E. J. Son, and H. M. Lee, Physical Review D **103**, 123023 (2021), arXiv:2101.05685 [astro-ph, physics:gr-qc].

[79] A. Khan, E. A. Huerta, and H. Zheng, Physical Review D **105**, 024024 (2022), arXiv:2110.06968 [astro-ph, physics:gr-qc].

[80] R. Shi, Y. Zhou, T. Zhao, Z. Cao, and Z. Ren, Physical Review D **109**, 084017 (2024).

[81] J. Yan, J. Luo, Y. Zeng, A. P. Leung, J. Feng, H.-H. Zhang, and W. Lin, The Astrophysical Journal **973**, 163 (2024).

[82] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Instance Normalization: The Missing Ingredient for Fast Stylization," (2017), arXiv:1607.08022.

[83] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," (2017), arxiv:1412.6980 [cs].

[84] A. Nitz, I. Harry, D. Brown, C. M. Biwer, J. Willis, T. D. Canton, C. Capano, T. Dent, L. Pekowsky, G. S. C. Davies, et al., "Gwastro/pycbc: V2.3.3 release of PyCBC," Zenodo (2024).

[85] T. B. Littenberg and N. J. Cornish, Physical Review D **80**, 063007 (2009).

[86] Z.-Q. J. Xu, Y. Zhang, T. Luo, Y. Xiao, and Z. Ma, Communications in Computational Physics **28**, 1746 (2020), arXiv:1901.06523 [cs, stat].

[87] Y. Bouffanais and E. K. Porter, Physical Review D **100**, 104023 (2019).