Deep generative models of gravitational waveforms via conditional autoencoder

Chung-Hao Liao[1,†] and Feng-Li Lin[1,2,*]

[1]*Department of Physics, National Taiwan Normal University, Taipei 11677, Taiwan*
[2]*Center of Astronomy and Gravitation, National Taiwan Normal University, Taipei 11677, Taiwan*

(Received 28 January 2021; accepted 1 June 2021; published 22 June 2021)

We construct few deep generative models of gravitational waveforms based on the semisupervising scheme of conditional autoencoders and its variational extensions. Once the training is done, we find that our best waveform model can generate the inspiral-merger waveforms of binary black hole coalescence with more than 97% average overlap matched filtering accuracy for the mass ratio between 1 and 10. Besides, the generation time of a single waveform takes about one millisecond, which is about 10 to 100 times faster than the effective-one-body-numerical-relativity algorithm running on the same computing facility. Moreover, these models can also help to explore the space of waveforms. That is, with mainly the low-mass-ratio training set, the resultant trained model is capable of generating large amount of accurate high-mass-ratio waveforms. This result implies that our generative model can speed up the waveform generation for the low latency search of gravitational wave events. With improvement of the accuracy in the future work, the generative waveform model may also help to speed up the parameter estimation and can assist the numerical relativity in generating the waveforms of higher mass ratio by progressively self-training.

DOI: 10.1103/PhysRevD.103.124051

I. INTRODUCTION

LIGO/Virgo has detected about a hundred of compact binary coalescence (CBC) up to its O3 observations [1–3]. This is remarkable achievement of modern science. Due to the limitation of LIGO/Virgo's sensitivity, these events are detected by the method of matched filtering [4–6], which calculates the overlap between the whitening data and the theoretical gravitational waveform templates. Similarly, the source properties of these events are also extracted based on matched filtering to perform the Markov chain Monte Carlo (MCMC) Bayesian parameter estimation (PE) [7,8]. In both processes of detection and PE of gravitational wave events, a huge number of theoretical waveform templates are required for matched filtering, therefore the efficiency of evaluating waveform templates is crucial for detection and to accelerate the PE procedures. However, due to the nonlinear feature of Einstein gravity and the unavoidable strong gravity regime for the mergers of two compact objects, it is notoriously difficult to calculate the CBC dynamics and the associated gravitational waveforms. For example, it is known [9,10] to require about 100000 CPU hours to obtain a state-of-art CBC waveform by solving numerical relativity. The required computing time will be increased by one or two orders for the higher mass-ratio CBC events, and is beyond what the current computing facility can afford. Thus, it is impractical to adopt such *ab initio* waveforms directly for performing either detection or PE.

To accelerate the generation of theoretical waveforms for practical applications, some analytical waveform models are introduced with a few parameters to be fitted by the results of numerical relativity. The well-known examples are IMRPhenomP models [11,12], the synergy models [13–15] that combine the post-Newtonian [16,17], effective-one-body (EOB) formalism [18,19], black hole perturbation [20,21], and numerical relativity [22,23], and the reduced order models or surrogate models [24–26] that span the generic waveforms with some orthonormal basis. However, it still takes a few hundredths to a few tenths of a second to evaluate a single waveform based on the aforementioned analytical waveform models.[1] By this speed of waveform generation, it will usually take weeks or even months to obtain the state-of-art PE results for a single event based on the MCMC algorithm. One can then expect the overall computing power cost or time span for PE will increase rapidly for the latter operations of LIGO/Virgo/KAGRA such as O4 or O5, for which the number of detection CBC events will be increased by an order or more. Therefore, the speeding-up of waveform generation becomes a pressing issue even in the near future.

[*]Corresponding author.
fengli.lin@gmail.com
[†]rossliao125@gmail.com

[1] This can be estimated by generating the waveforms from template library in either PyCBC or GstLAL.

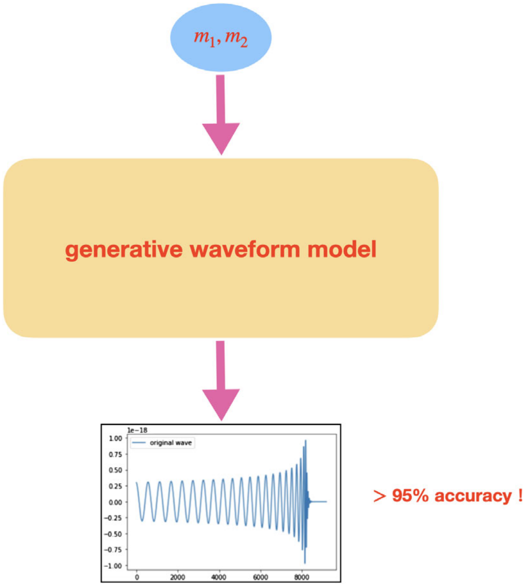

FIG. 1. A generative model of gravitational waveforms. Once the neural network model is well trained, it can generate the gravitational waveforms with more than 95% accuracy when providing just the source labels such masses (m_1, m_2) of the binary black holes. The accuracy rate of the waveforms is defined in (8) below. As a preliminary study for the proof of concept, in this work we mainly consider the inspiral-merger parts of the full waveforms.

Besides, those aforementioned analytical waveform models are in nature interpolating models by fitting the parameters with a known set of waveforms. This implies that the model could become more complicated and cumbersome when the range of the waveforms are extended, such as going to a higher mass ratio. The increase of the complexity will reduce the models' efficiency of generating the real-time waveforms for the detection or PE. Thus, it is crucial to have some extrapolating models of waveform generation to resolve the conflict between complexity and efficiency of the traditional analytical waveform models.

Motivated by the above discussions on the limitations of the known models of waveform generation, we turns to the deep learning for the resolution. We aim to construct some deep learning neural network to generate the CBC gravitational waveforms of high accuracy by giving the source parameters such as the masses, spins of the binary compact objects, as schematically depicted in Fig. 1. Even the training time will be increased as the training set is enlarged, the time of evaluating a new waveform with the trained machine will not be increased much. This then resolves the conflict between the complexity and efficiency for the real-time applications.

Moreover, we also hope this deep learning neural network to be generative so that it can generate the waveforms that do not belong to the source parameter ranges of the training set. For example, we can train the machine with waveforms of only low mass ratio (LMR), and then generate the accurate waveforms of higher mass ratio (HMR). This could help to efficiently obtain the HMR waveforms, which will be computationally costly by numerical relativity. However, in this work we will not explore this scenario but just a toy version, for which we employ the training set containing small fraction of HMR waveforms to demonstrate the possibility.

In view of the above target features, this deep learning machine should be the supervised one when training with the given source parameters and the associated waveforms. On the other hand, it is also better to be generative and the unsupervised one so that it has the potential to turn into an extrapolating model of generating the HMR waveforms. For this purpose, in this paper we adopt the conditional (variational) autoencoder (CAE or CVAE) [27–29] to construct various deep learning models to generate CBC gravitational waveforms.[2] This scheme belongs to the so-called semisupervised learning by combining both features of supervised and unsupervised learning.[3] It is built on a more basic scheme for the unsupervised learning, the autoencoder (AE) [35], or its generative extension, the variational autoencoder (VAE) [36,37]. We will introduce the basics of these neural networks in the next section.

As a preliminary study for the proof-of-concept, in this work we mainly consider the inspiral-merger parts of the full waveforms but truncating the ringdown part. We find that our best generative models can produce the waveforms with accuracy higher than 97% even for the generation of HMR waveforms. Moreover, it can produce a single waveform within one millisecond, which is about 10 to 100 times faster than producing an EOB waveform on the same computing facility. To visualize the accuracy rate, in Fig. 2 we shows some typical examples of the waveforms with different accuracy rates.

The rest of the paper is organized as follows. In the next section, we will briefly sketch the basics of autoencoder and its extensions including VAE and the conditional versions. In Sec. III we describe the tomography of our training data set, and how we prepare our training waveforms. Besides, the fitting factor or faithfulness (FF) based on the overlap of matched filtering is introduced to characterize the accuracy of the generative waveform models. In Sec. IV we consider four waveform models based on the CAE scheme, and then summarize their accuracy and run-time in Tables II and IV, respectively. By comparing the accuracy, we pick up the best CAE waveform model and present its detailed information. Finally, we conclude this paper in Sec. V. In the Appendix, we present the performance of the CVAE counterparts of the CAE waveform models considered in the main text.

[2] The CVAE framework is recently adopted as the generative model of posteriors for the PE of the CBC events, see [30–32].

[3] For the basic discussion of supervised and unsupervised learning, please see [33,34].

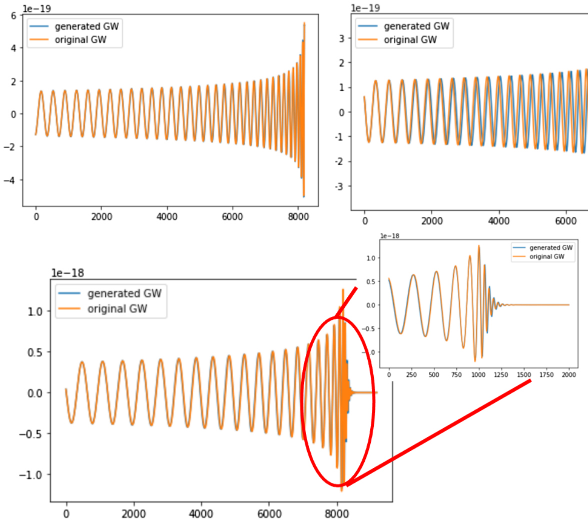

FIG. 2. Some typical examples of the generated waveforms by well-trained CVAE models with different accuracy rates. Top left: a generated inspiral-merger waveform of 99.50% accuracy when comparing with the corresponding EOB waveform by overlap match. Top right: a generated inspiral-merger waveform of 92.37% accuracy. Bottom: a generated full inspiral-merger-ringdown waveform of 99.74% accuracy.

II. AUTOENCODER AND ITS EXTENSIONS

Our goal is to construct some deep learning models of gravitational waveforms as depicted in Fig. 1. The basic structure of this generative model is the so-called AE [35] or it extension, the VAE [36,37]. The basic structure of AE and VAE is shown in Fig. 3, which contains two parts: the encoder and the decoder. The encoder [denoted by $q_\phi(z|x)$ with ϕ the abbreviation of biases and weights of the encoder's neural network] compresses the input data x into the latent layer z of smaller dimensions than the ones of x, and then the decoder (denoted by $p_\theta(\tilde{x}|z)$ with θ the abbreviation of biases and weights of the decoder's neural network) uncompresses the latent layer back to the final result $\tilde{x}$ of the same dimension as x. One then use some distance measure such as mean-squared error (MSE) between x and $\tilde{x}$ as the reconstruction loss. The goal is to minimize the reconstruction loss to optimize the biases and weights of the whole AE's neural network. Since there is no label for the input data, this is the unsupervised learning.

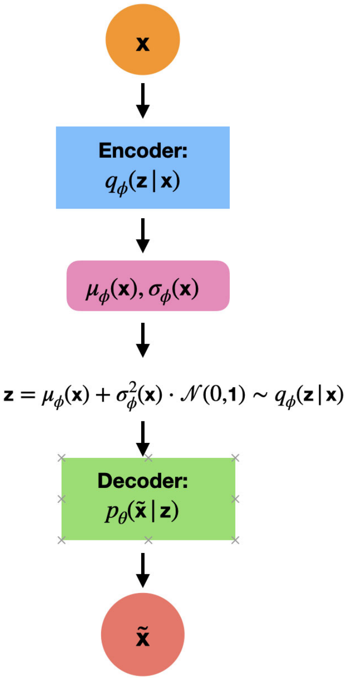
x
Encoder:
$q_\phi(\mathbf{z}\,|\,\mathbf{x})$
$\mu_\phi(\mathbf{x}), \sigma_\phi(\mathbf{x})$
$\mathbf{z} = \mu_\phi(\mathbf{x}) + \sigma^2_\phi(\mathbf{x}) \cdot \mathcal{N}(0,1) \sim q_\phi(\mathbf{z}\,|\,\mathbf{x})$
Decoder:
$p_\theta(\tilde{\mathbf{x}}\,|\,\mathbf{z})$
$\tilde{\mathbf{x}}$

FIG. 3. Schematic structure of a AE or VAE. It contains two components. (i) An encoder $q_\phi(z|x)$ which transforms an input vector x to a latent vector z, which is deterministic for AE, but stochastic for VAE, i.e., $z = \mu_\phi(x) + \sigma^2_\phi(x)\mathcal{N}(0,1)$. Here $\mathcal{N}(0,1)$ is the unit normal distribution. (ii) An decoder $p_\theta(\tilde{x}|z)$ that transforms x to an output $\tilde{x}$. The loss function of AE is just the reconstruction loss such as a MSE between $\tilde{x}$ and x. On the other hand, the loss function of VAE contains two parts: the reconstruction loss and the KL loss as discussed in (1).

Since the AE is a deterministic machine so that it may lack the power of extrapolations and could fail to be generative. To remedy this drawback, the VAE is introduced by making the latent layer a stochastic one. This is done by generating the means and variances of the Gaussian distributions as the output of the encoder, from which one can sample a latent layer as the input to the decoder, as shown in the middle of Fig. 3. The uncertainty of the layer make the VAE to be able to "think out of the box," and is thus a generative machine. However, besides the reconstruction loss one should also consider the regularization loss which characterizes how much the stochastic latent layer deviates from $\mathcal{N}(0,1)$, i.e., the unit Gaussian with zero mean. This is measured by their Kullbac-Leibler (KL) divergence. It turns out that the combined loss is equal to upper bound of the negative of the log likelihood of the input data distribution $p_\theta(x)$, i.e.,

$$-\log p_\theta(x) \le \mathbf{E}_{z\sim q_\phi(z|x)}[-\log p_\theta(\tilde{x}|z)] + \mathbf{D}_{\mathrm{KL}}[q_\phi(z|x)||\mathcal{N}(0,1)], \tag{1}$$

where the first term on the right-hand side is the reconstruction loss and the second term is the regularization loss.

When training the waveform models, the input x from the training dataset is the theoretical waveform such as EOB waveform, and we call it strain for short. We can choose the reconstruction loss to be the MSE between x and $\tilde{x}$. The training process is to optimize the biases and weights of the neural network by minimizing the reconstruction loss

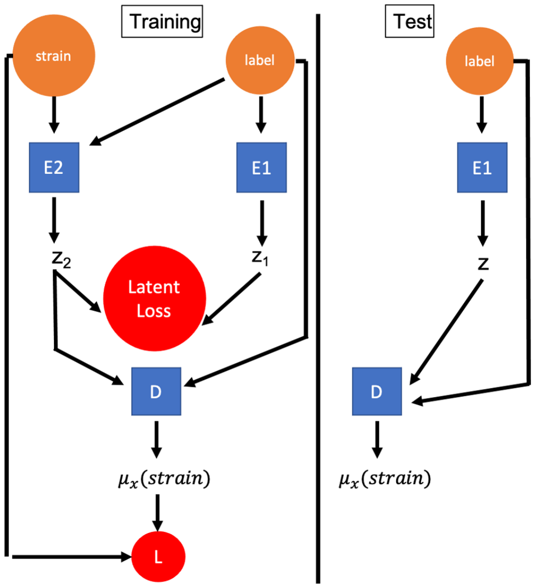

FIG. 4. The schematic structure of a CAE or CVAE as a generative waveform model. Left panel: during the training period, it needs two encoders: (a) one for training the input data such as strains/waveforms, and (b) one for training the source labels associated with the input data such as (m_1, m_2). Right panel: after the training, the encoder (a) is removed so that it becomes a generative model, namely it generates waveforms by providing only the associated source labels.

such that the generated $\tilde{x}$ can be as close to x as possible. After the training, the decoder can be turned into a generative model of strains, namely, given some input latent vector, the decoder will output some strain. However, this machine is not so useful in generating the strains with specific source properties because the latent space may not correspond to the required parameter space of the physical source properties, such as masses (m_1, m_2) of the binary black holes. For convenience we call the source parameters the labels. To make the AE or VAE useful for our purpose, we adopt the way of semisupervised training by also conditioning the labels when training the machine. After the training we will truncate the encoder part associated with the strain input, the remaining one with the label as input will then become the useful generative model of strains, namely, given the label such as (m_1, m_2), the machine will generate the associated strain. The above scheme is called the conditional AE/VAE abbreviated as CAE/CVAE [27–29], and the basic structure is depicted in Fig. 4 where an additional encoder for the labels of input data is introduced.

Due to the additional encoder, we now have two latent vectors z_1 and z_2 as shown in Fig. 4. We can then introduce the latent loss to measure their difference. For AE, the latent loss can be MSE between latent vectors, but for VAE it is the KL divergence between the Gaussian distributions generated by the two encoders. On the other hand, the reconstruction loss is the MSE for both AE or VAE.

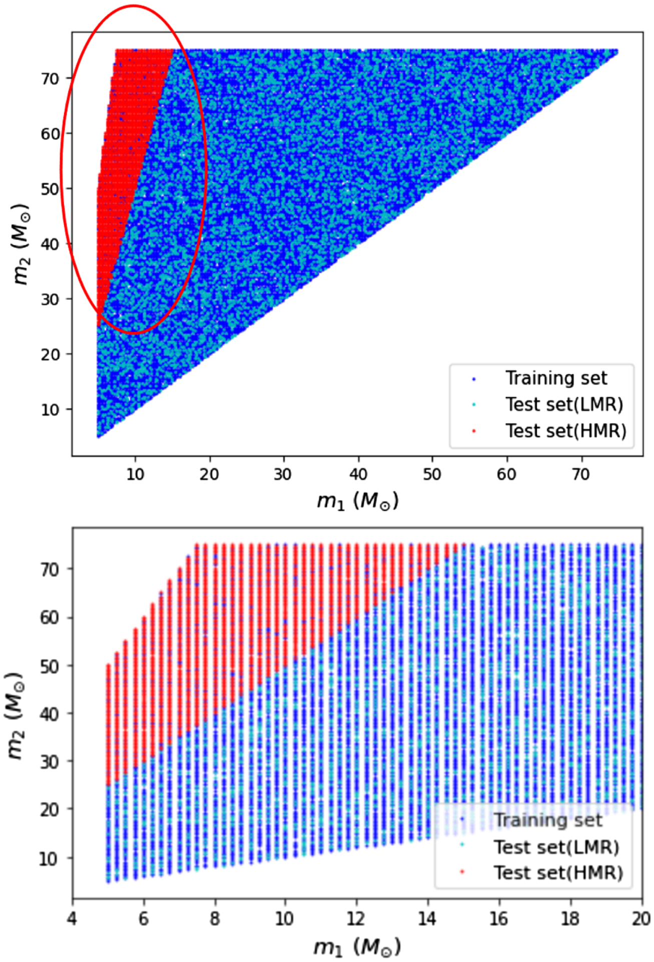

FIG. 5. Tomography of dataset for training, validation and test of a CAE model with the numeric as listed in Table I. Top: overview of the tomography. Bottom: enlarged view of some portion circled in the top figure, with more clear visual estimate.

III. WAVEFORM DATA PREPARATION AND OVERLAP ACCURACY

Once we construct the code for CAE or CVAE, we prepare a set of strains to train the machine. A strain is the linear combination of the two polarization modes $h_+(t)$ and $h_\times(t)$, i.e.,

$$h(t) = h_+(t) + ih_\times(t). \tag{2}$$

To be specific, we consider the inspiral-merger part of the CBC strains of binary black holes with their masses (m_1, m_2) as the only source parameters, i.e., without spin and procession. We further divide the set into two subsets, one is called the LMR set for $q = m_2/m_1 \leq 5$ and the other one called the HMR set for $q > 5$. These strains are obtained from ENOBNR [14] of the PyCBC library [7], in which each strain is divided into 8192 time segments. We basically train the machine mainly with the LMR set but combining with about 20% of HMR strains. The latter is

TABLE I. The range of source parameters and the amounts of the data set. Here the mass ratio is denoted by $q \equiv m_2/m_1$ with $1 \le q \le 10$. The corresponding percentages of (training, validation, test) is (70%, 10%, 20%) for LMR (low-mass-ratio) data ($q \le 5$), and is (19%, 1%, 80%) for HMR (high-mass-ratio) data ($q > 5$). Note that the fraction of the HMR templates is only about 2.46% of the total training data set, including both training and validation data.

	m_1	m_2	q	Δm	Train	Valid	Test
$q \le 5$	[5.0,75.0]	[5.0,75.0]	[1,5]	0.25	24865	3552	7104
$q > 5$	[5.0,75.0]	[5.0,75.0]	[5,10]	0.25	682	36	2873

used as the tutor seed to train the machine toward a generative model for the other 80% of HMR strains. The tomography of our training and test sets is shown in Fig. 5, and in Table I we give the more details for the specs of this tomography. Moreover, the fraction of the HMR templates in the total training set is only about 2.46%. This tiny fraction is chosen on purpose to mimic as closely as possible the real extrapolating model for which the training set contains no HMR waveform.

As a preliminary study to demonstrate that a generative model of gravitational waveform is in principle possible, we will not consider the full CBC strain but truncate the ringdown part, which is far shorter than the other part of the strain. The truncated waveform is denoted as the inspiral-merger strain. The purpose of this truncation is to further reduce the complexity of the frequency/amplitude part of a strain caused by the sudden change at the merger, and will help to well train the machine with less efforts in tuning the hyperparameters.

Using the time series form of the inspiral-merger strains to train a CAE or CVAE, the result turns out to be not good for reasonable machine size and training, see Fig. 6 for a typical result. It implies that the model cannot catch up the amplitude and phase correctly at the same time. This suggests that this form of strain is still too complicated for a CAE or CVAE of reasonable size to work properly. Motivated by this result, we then decide to separate the amplitude and frequency parts of a strain, and then juxtapose them as the input of the CAE or CVAE. To be specific, from the two polarization modes we first obtain the instantaneous phase

$$\theta(t) = \tan^{-1}\left(\frac{h_\times(t)}{h_+(t)}\right), \qquad (3)$$

and the instantaneous frequency and amplitude are given, respectively, by

$$\omega\left(t + \frac{\delta t}{2}\right) = \frac{\theta(t+\delta t) - \theta(t)}{2\pi\delta t}, \qquad (4)$$

$$A(t) = \sqrt{h_+^2(t) + h_\times^2(t)}. \qquad (5)$$

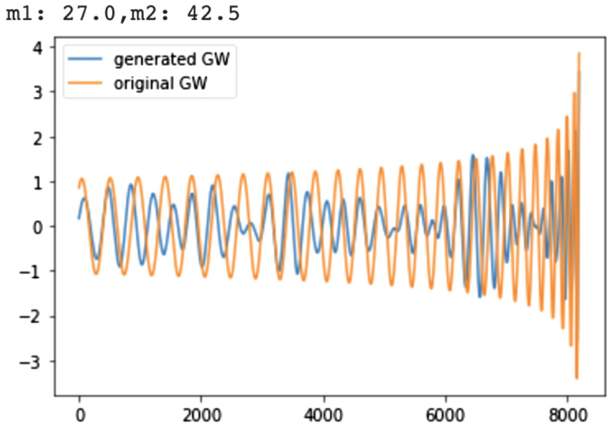

FIG. 6. A typical generative waveform (blue color) from a trained CAE or CVAE by inputting a inspiral-merger parts of the CBC strains (orange color) in a time series format. It cannot catch both the phase and amplitude at the same time.

A typical example showing the above decomposition is given in Fig. 7.

Even using this frequency/amplitude separated form of the strains to train the CAE or CVAE model, the result is still not good because the magnitudes of the input data have not been rescaled to avoid too small or too large values. This however can be solved as in the usual deep learning process for neural network by just normalizing the input data [38]. The way of normalization we adopt is as follows:

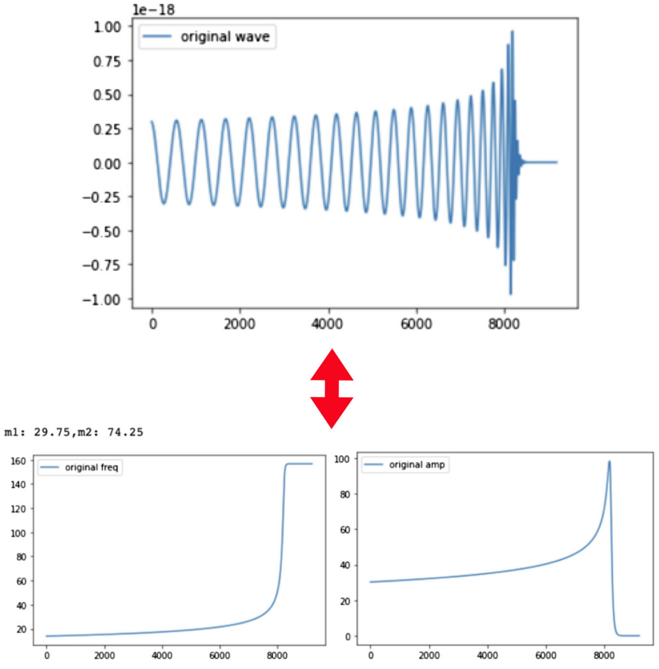

FIG. 7. Decomposition of a time series strain $h(t)$ (top) into frequency (bottom left) and amplitude (bottom right) by using (4) and (5). The amplitude part has been multiplied by a factor of 10^{20} to have comparable scale with the frequency one.

$$\hat{\omega}(t) = \frac{\omega(t) - \mu_\omega}{\sigma_\omega}, \qquad \hat{A}(t) = \frac{A(t) - \mu_A}{\sigma_A}, \tag{6}$$

where the normalization parameters $(\mu_\omega, \sigma_\omega)$ are, respectively, mean and variance evaluated from the 8192 segments of $\omega(t)$,[4] and similarly for (μ_A, σ_A). We call these four parameters the key (to reconstruct the associated unnormalized strain).

Naively, we can juxtapose these four normalization parameters with the normalized strain vector $(\hat{\omega}(t), \hat{A}(t))$ of 2×8192 segments to form the input of CAE or CVAE. However, their dimensions are not in proportion, the juxtaposition could suppress the significance of the key during the training, which will induce the unbearable error in recovering the full strain via (6). Therefore, we need to find the appropriate CAE or CVAE schemes to well train both the normalizing strains and the associated keys to the desirable accuracy.

Once the training of a waveform deep learning model is done, we need to evaluate their performance based on some criterion of accuracy by comparing a machine-generated waveforms $h_{\rm ML}(t)$ with the corresponding waveform $h_{\rm EOB}(t)$ obtained from effective-one-body-numerical-relativity (EOBNR). To calculate the accuracy we adopt the conventional overlap method used in gravitational waveform community. The overlap method is motivated by the matched filtering [4–6] for the signal detection or parameter estimation, in which the overlap between two waveforms $h_1(t)$ and $h_2(t)$ is defined by

$$\langle h_1 | h_2 \rangle = 4{\rm Re} \int_0^\infty \frac{\tilde{h}_1(f) \tilde{h}_2(f)^*}{S_n(f)} df, \tag{7}$$

where $\tilde{h}_i(f)$ is the Fourier transform of $h_i(t)$ and $S_n(f)$ is the power spectral density of the detector's noise. In practical, some appropriate low and high frequency cutoffs will be imposed when performing the integral. To evaluate the accuracy of a waveform model, the following FF or faithfulness [14,26,39] is adopted to compare $h_{\rm ML}(t)$ generated by our waveform model and the standard EOB waveform $h_{\rm EOB}(t)$,

$$\mathrm{FF} = \max_{t_0, \phi_0} \left[\frac{\langle h_{\rm EOB} | h_{\rm ML} \rangle}{\sqrt{\langle h_{\rm EOB} | h_{\rm EOB} \rangle \langle h_{\rm ML} | h_{\rm ML} \rangle}} \right], \tag{8}$$

where t_0 and ϕ_0 are, respectively, the initial time and inital phase of $h_{\rm EOB}(t)$. Without being biased by the detector noise, below we will choose flat power spectral density, i.e., $S_n(f) = 1$ for evaluating the FF [26]. To characterize the performance, we need to evaluate the FF of each template in the test dataset, i.e., 20% of LMR and 80% of HMR, and find out the distribution of FFs, which can also be represented by its maximum, median, and minimum. However, for simplicity we can represent and denote the accuracy simply by the average of the FFs over the test dataset. This may not be precise enough but is more convenient when comparing the performances of different generative waveform model. Later on, we will give the cumulative distribution function of FFs and the associated maximum, median, and minimum for the best model selected by comparing the average of FFs over the test dataset. Moreover, at current stage the initial phase is not optimized when evaluating FF. Despite that, our best waveform models can be shown to achieve more than 97% accuracy even without optimizing the initial phase. Once the initial phase is also optimized, the accuracy can be expected to be further enhanced.

Note that it turns out that the CVAE models yield comparable but lower accuracy than the CAE models. This is probably because the template datasets considered in this paper are parametrized by two mass parameters, which do not cause much degeneracy in mapping the parameter space to the template space. The degeneracy here means that the different set of parameters may yield quite similar waveforms. Thus, the variational feature of latent space of CVAE models may not be needed for such kind of deterministic training set. Despite that, when considering more complicated template sets with more source parameters, the variational feature could be helpful to disentangle the degeneracy, which may occur more often. In this respect, it is still interesting to consider the VAE type models as a preliminary study for the future work. To not digress the main theme of this work, from now on we will simply focus on the various models based on the CAE scheme in the main text. As a comparison, in the Appendix we present the performance of CVAE models with the same schematic structures of the CAE models.

IV. CONDITIONAL AUTOENCODER WAVEFORM MODELS

Based on the CAE scheme we can construct various waveform models by different arrangements of the encoders and decoders. Since the input data are separated into keys and normalized strains, their associated CAE can be arranged to share a common decoder or not. For either cases, we consider two waveform models, which further differ by how the labels are conditioned. In total, we will consider four waveform models and compare their performances. In this way, we can understand the relevance of different arrangements to the performance, so that such experiences could be helpful for further constructions. Below we first consider the models in which the keys and normalized strains do not share a common decoder, and then the models do.

The first CAE waveform model as shown in Fig. 8 is what we call CAE + NN model, in which the strains are trained with CAE and the associated keys are trained with conventional supervised learning neural network (NN)

[4] Due to the nature of its definition from the difference between two neighbor segments, there are only 8191 segments for $\omega(t)$.

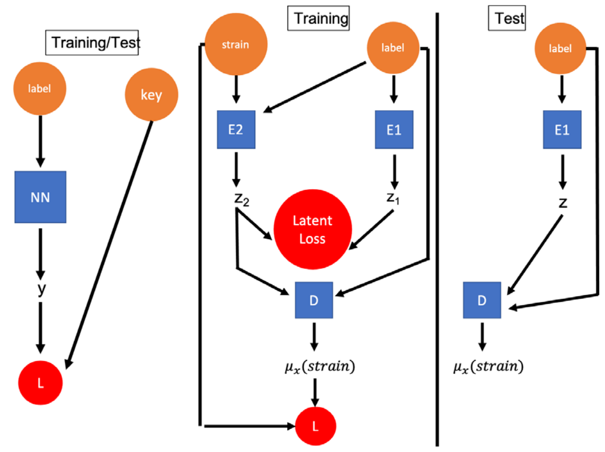

FIG. 8. Schematic structure of the CAE + NN waveform model.

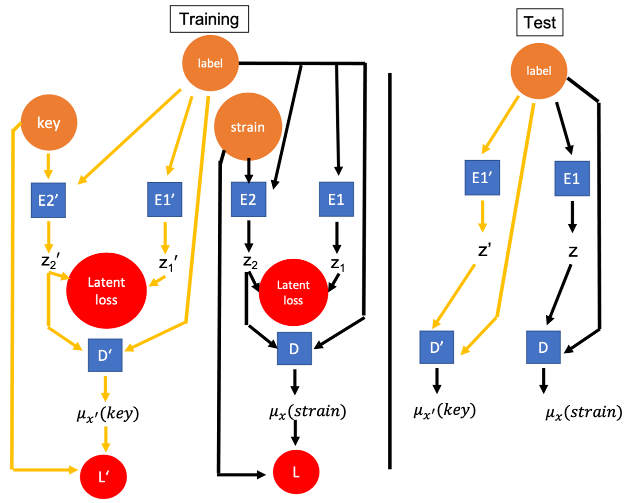

FIG. 9. Schematic structure of the 2CAE waveform model.

because the dimensions of the key are relatively small. After done the training, we can drop the strain part from the model, and turn the remaining network (right-most part of Fig. 8) into a generative machine for waveforms. Since the keys and the normalized amplitude and frequency parts of the strains are trained separately, when generating a strain we need to combine them together to get the un-normalized amplitude $A(t)$ and frequency $\omega(t)$. Finally, we need to integrate the frequency to get the phase $\theta(t)$ and then combine with the amplitude to get the strain $h_{\mathrm{ML}}(t)$. With the output strain $h_{\mathrm{ML}}(t)$ we can use (8) to evaluate the FF for each waveform in test dataset, and then take the average to obtain the accuracy. The resultant performance is 85.73% for LMR, and 55.95% for HMR. As expected, the accuracy for HMR is lower than the one for LMR. The accuracy is not good enough for the purpose of data analysis. This is because the simple NN for training the key is not the optimal scheme as AE.

By the faith on the power of semisupervised training, we can replace the NN by AE to train the key, and the new scheme is shown in Fig. 9. As expected, the performance for both LMR and HMR get improved. The resultant accuracy is 89.92% for LMR and 67.20% for HMR. The accuracy of LMR is now barely good for detection purpose, but not good for PE to extract the source properties. Still the HRM part is not accurate enough for practical purpose.

A common feature of the above two models is that they train the keys and strains separately, and the correlation is only through the common labels. Instead, we can correlate the outputs of the different encoders into a common decoder, and directly compare the decoder's output to the corresponding un-normalized frequency and amplitude of the input strain to obtain the reconstruction loss. Intuitively, the additional correlation may improve the accuracy of the generative model. We will consider two such kinds of models. The first one is shown in Fig. 10, which we call 1C2E1D, i.e., one conditional encoder, and two waveform encoders (one for the strain and one for the key), and finally one common decoder. In this model, the latent sizes of all the encoders should be the same. This may cause redundancy of the latent space for training the key since the key's dimension is far less than the strain's. To remedy this, we introduce the second model as shown in Fig. 11. We call this model 2C2E1D, i.e., now we have two conditional encoders so that the latent sizes of the encoders for the strain and the key can be different. Moreover, since now the keys and the normalized strains share a common encoder, we can in fact choose the target in Figs. 10 and 11 to be the un-normalized amplitude and frequency, i.e., $A(t)$ and $\omega(t)$, but not the normalized $\hat{A}(t)$ and $\hat{\omega}(t)$, to evaluate the reconstruction loss. For simplicity, we choose the MSE for the reconstruction loss so that the minimization of the reconstruction loss will make the machine to generate $\tilde{x}$

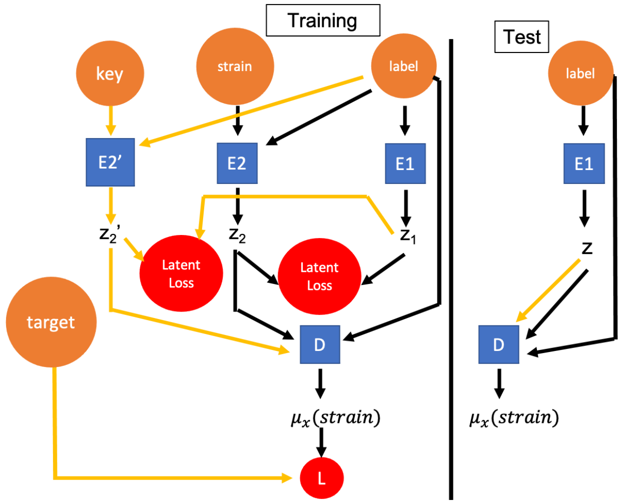

FIG. 10. Schematic structure of the 1C2E1D waveform model.

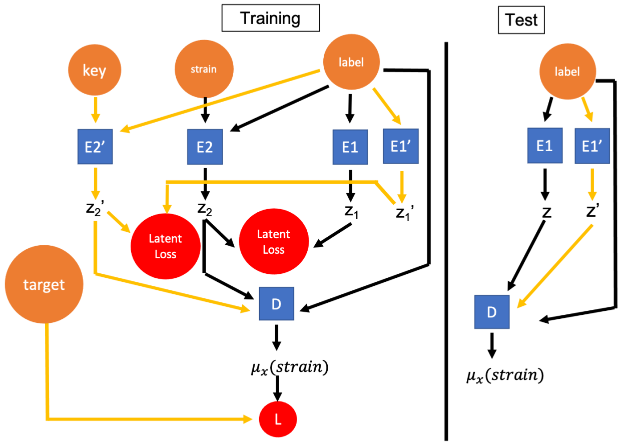

FIG. 11. Schematic structure of the 2C2E1D waveform model.

TABLE II. Summary of the accuracy of the LMR and HMR waveforms for each CAE waveform model considered in this paper. The accuracy is the average of the FFs [see (8)] for all the test data. We see that both 1C2E1D and 2C2E1D models can have accuracy more than 97% for both LMR and HMR.

	CAE + NN	2CAE	1C2E1D	2C2E1D
Accuracy (LMR)	85.73%	89.92%	97.65%	98.20%
Accuracy (HMR)	55.95%	67.20%	97.70%	97.02%

TABLE III. Summary of the FFs of the LMR and HMR waveforms for 2C2E1D CAE waveform model considered in this paper. The associated cumulative distribution function of FFs is shown in Fig. 14. We see that the medians are comparable with accuracy listed in Table II.

FFs for 2C2E1D	LMR	HMR
Minimum FF	82.49%	74.13%
Median FF	98.61%	98.22%
Maximum FF	100.0%	99.99%

being as close to the target $A(t)$ and $\omega(t)$ as possible. In contrast to the CAE + NN and 2CAE waveform models, this will save the procedure of converting the normalized strains into their un-normalized counterparts.

The resultant performance for the above two models are the following. For the 1C2E1D waveform model, the accuracy is 97.65% for LMR, and 97.70% for HMR. For the 2C2E1D waveform model, the accuracy is 98.20% for LMR, and 97.02% for HMR. Their performances are comparable, and are accurate enough for both LMR and HMR, i.e., greater than 97%. Note that we have not optimized the overlap accuracy over the initial phase yet, once it is done we will expect higher overlap accuracy.[5] The high accuracy of the generated waveforms indicates that both models are good for the purpose of low latency detection and for PE of gravitational wave events with improvement of accuracy in the future work. The high accuracy for HMR part can also be exploited for progressively self-training to generate waveforms of HMR.

We summarize in Table II the accuracy for each CAE waveform model considered in this paper. The 1C2E1D and 2C2E1D models are the best in the accuracy rate. Overall the 2C2E1D model is slightly superior. To characterize its detailed performance, we also give the minimum FF, median FF, and maximum FF of this model in Table III, from which we see the median FF is comparable with the accuracy, i.e., the average of FFs. Later we will discuss more details for this model. Note that, the above models are all implemented based on the CAE scheme. We can also replace the CAE schemes in these models by the CVAE ones, and obtain the corresponding CVAE waveform models. However, there is one additional model called CVAE + CAE, see Fig. 15 in the Appendix, in which we use CAE to train the keys and CVAE to train the strains. The performances of these CVAE waveform models are listed in the Table VI of the Appendix. It turns out that the accuracy of the CVAE waveform models are comparable to their CAE counterparts. However, by examining in more details it seems that CAE models are superior than the CVAE ones, even for the HMR. This is a bit surprising that the generative nature of VAE does not help to improve the accuracy.

Besides, we also summarize in Table IV the training time and generation/epoch number of the waveform models and the generation time of a single waveform for each CAE waveform model considered in this paper. The run-times of the CVAE waveform models are comparable with their

TABLE IV. Summary of the training time, generation/epoch number and the generation time of a single waveform for each CAE waveform model considered in this paper. Note that it takes less than 1 millisecond to generate a single waveform. This is about 10 to 100 times faster than the EOB running on the same computing facility.

	CAE + NN	2CAE	1C2E1D (CAE)	2C2E1D (CAE)
Training time (strain)(sec)	4042.5	3329.4	4536.6	4462.6
Training time (key)(sec)	81.2	144.8		
Generations/epochs (strain)	8000	8000	10000	10000
Generations/epochs (key)	10000	10000		
Generation time per waveform (milli sec)	0.8–1.0			

[5] Our preliminary study shows that it can achieve almost 99% for LMR and 98% for HMR.

TABLE V. Hyperparameters for the 2C2E1D model. Here “conv features” is the abbreviation of the convolution features.

	E_1	E_2	E_1'	E_2'	Decoder
Latent size	8	8	3	3	8&3
CNN layers	None	2	None	None	3
Filter size	None	[16,16]	None	None	[4,4,4]
conv features	None	[5,15]	None	None	[16,32,64]
Pool size	None	[4,4]	None	None	[4,4,4]
Dilation rate	None	None	None	None	[1,2,2]
NN layers	4	3	4	3	6
Neural size	500	500	400	400	800

CAE counterparts listed in Table IV, and thus are omitted for simplicity. We see that the training time is about 4000 seconds for all the waveform models, it is quite modest and implies that the extension to the full waveform models with more source parameters is manageable in the near future. Furthermore, the generation time of a single waveform is about one millisecond. Compared to the typical generation time for a EOB waveform, which is about few hundredths to few tenths of a second on the same computing facility, the speed enhancement is about 10 to 100 times.

As the 2C2E1D waveform model is the best among all the waveform models considered in this paper, we look into some details of this model. First, we list the hyperparameters of this model in in Table V, and histogram of its training losses in Fig. 12. Based on the information one can reproduce the model quite easily. From Fig. 12 we see that the training and validation losses match well and stop increasing around 8000 generations/epochs. This implies that our training is not overfitted and stabilized.

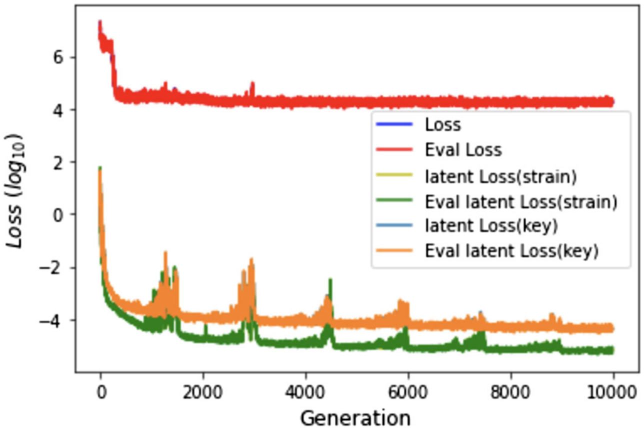

FIG. 12. Histogram of the training losses for the 2C2E1D CAE waveform model. The types of training/validation losses are denoted by [Loss/Eval_loss, latent Loss(strain)/Eval latent Loss (strain), latent Loss(key)/Eval latent Loss(key)] in the graphic illustrations, which literally mean the total loss, the latent loss of the normalized strain, and the latent loss of the key, respectively. The match of Loss and latent Loss implies no overfitting. The overall trends show that the training is stabilized around 10000 generations/epochs.

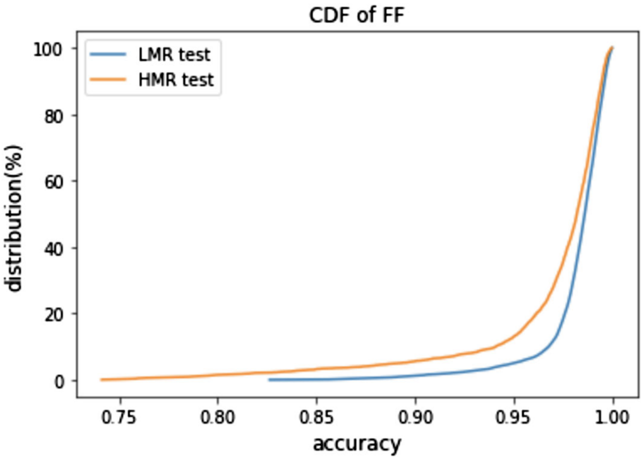

FIG. 13. Cumulative distributions functions of FFs of 2C2E1D CAE waveform model for the LMR (blue) and HMR (orange) generated waveforms. As expected, the HMR one has a broader tail. Overall, the outliers are rare.

Further, we can understand more the tomography of the accuracy for the 2C2E1D CAE waveform model by plotting the cumulative distributions function of the FF for both LMR and HMR. The results are shown in Fig. 13. As expected, we see that the HMR one has a broader tail; however, overall the FFs are concentrated on the side of high FF near 100%. This implies the outliers are rare, and the generated waveforms can be reliably implemented for the practical data analysis such as the detection and PE of the gravitational wave events. For curiosity, it is also interesting to see how FF changes with the mass ratio and total mass of the binary black hole. We present the distributions for 2C2E1D CAE waveform model in Fig. 14. We see that the low FFs appear more often in the regime of

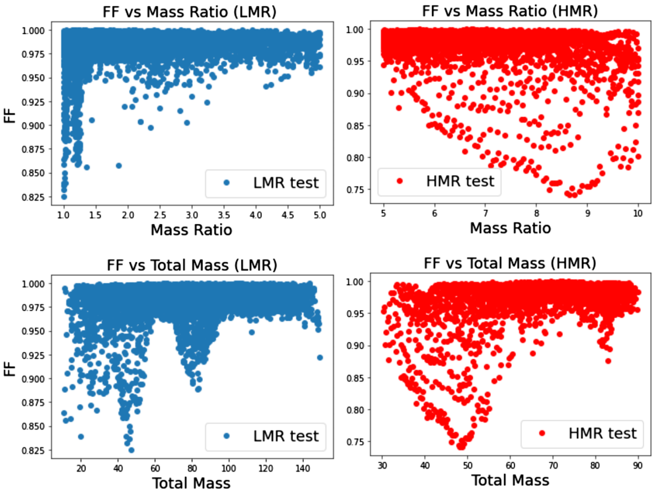

FIG. 14. Distributions of FFs of 2c1E1D CAE waveform model as functions of mass ratio (upper row) and total mass (lower row) for LMR (blue) and HMR (red) test dataset. Note that each dot represents one template in the test dataset.

lower mass ratio and smaller total mass for LMR ones. The latter one could be due to the fact that the templates in this regime are less loud (small mass), but it is not clear about the cause for the former one (lower mass ratio). On the other hand, the FF for the HMR does not behave the same way, which could be due to the insufficient training data.

Finally, the typical neural network structure of the above CAE models is given in Fig. 16 in the Appendix.

V. CONCLUSION AND DISCUSSION

In this paper we construct various waveform models based on the neural network of CAE and its variational extensions (CVAE). Their accuracy and run-time have been summarized in Tables II and IV in the main text for CAE models, and in Table VI in Appendix for CVAE models, respectively. For simplicity, we represent the accuracy of these generative waveform models by the average of fitting factor, which is based on the waveform overlap of the matched filtering. Among these waveform models, the so-called 2C2E1D CAE model is the best with more than 97% for both the LMR and HMR waveform generation. This demonstrates the viability of our best waveform model to be implemented in the practical gravitational wave data analysis and PE. Especially, the generation time of a single waveform is 10 to 100 times faster than the traditional EOBNR method, it implies that the waveform generation for the low latency detection can be accelerated by our waveform models. With the improvement of the accuracy in the future work, the revised version of our generative waveform model may also help to speed the parameter estimation.

Moreover, the impressive accuracy for HMR waveform generations is encouraging because fraction of the HMR waveforms in the training and validation dataset is less than 3%. This implies that one may be able to generate higher mass-ratio waveforms by a series of self-training with the generative outputs of the lower-mass ratio machine as the training data for the higher mass-ratio ones. This may open a new venue to generate the waveforms with intermediate mass-ratio, say greater than 15.

Despite that, there are still ample space to improve our waveform models. As a proof-of-concept study, we only consider the inspiral-merger part of the full waveforms. Although the ringdown part is quite short, it contains the information of quasinormal modes. We are currently training the waveform models for the full waveform based on the similar CAE or CVAE schemes, and will report our results in the near future. Moreover, to be more useful in the practical data analysis tasks, we shall also include more source parameters such as spins, precession, and tidal deformabilities. Once the above goals are achieved, we can incorporate our waveform models to the standard pipeline of detection and PE, and help to accelerate the data analysis tasks in the coming O4 and O5 observation runs of LIGO/Virgo/KAGRA.

Note.—Recently, an eprint [40] with the similar goal appears, in which a recurrent neural networks framework is adopted to generate the merger-ringdown parts of the waveform from the input associated inspiral one.

ACKNOWLEDGMENTS

We thank Kai-Feng Chen, Yao-Yu (Joshua) Lin, and members of TGWG and CAG for discussions and comments. In particular, we also thank Han-Shiang Kuo for the discussions on CVAE structures, and Jie-Shiun Tsao for helping the sever implementation. C. H. L. would like to thank Wei-Ren Xu for his guidance into the machine learning. This work is supported by Taiwan's Ministry of Science and Technology (MoST) through Grant No. 109-2112-M-003-007-MY3. We also thank NCTS for partial financial support.

APPENDIX: STRUCTURE AND PERFORMANCE OF CVAE WAVEFORM GENERATORS

In this appendix, we summarize the performance of the CVAE counterpart of the CAE waveform models considered in the main text. These counterpart models are simply obtained by replacing the CAE with CVAE in the associated CAE waveform model. However, for the 2CAE model, we can in fact replace only the CAE for the strains by CVAE, and still keep the CAE for the keys intact. In this way, we have a new model called CVAE + CAE model as shown in Fig. 15. For all the CAE and CVAE waveforms models considered in this work, the typical neural network structure is shown in Fig. 16, which serve as the guideline for the readers to implement the coding.

The accuracy for the CVAE waveform models are shown in Table VI, from which one can compare with Table II and finds that the accuracy are comparable for the CVAE models and their counterparts. Besides, the run-times for these CVAE models are comparable with their CAE

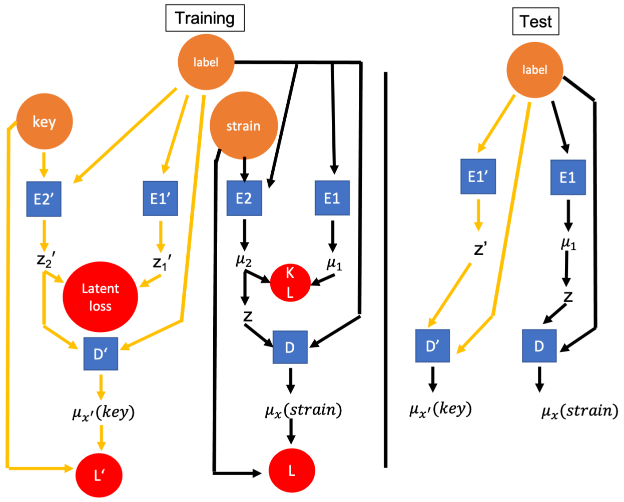

FIG. 15. Schematic structure of the CVAE + CAE waveform model.

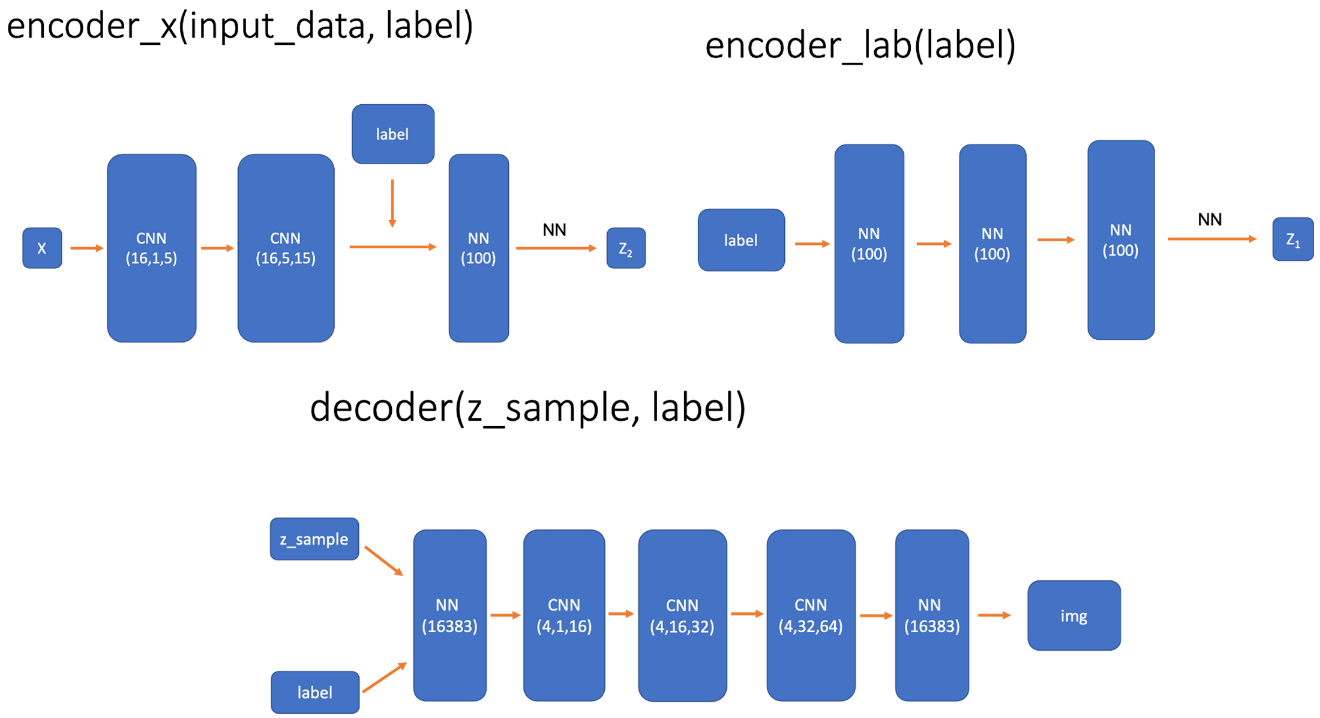

FIG. 16. A typical machine structure with details of hyperparameters for the CAE or CVAE used in this work. Each NN or convolutional neural network (CNN) is denoted by a box with its dimension specified. Top left: the encoder for the strain. Top right: the encoder for the label. Bottom: the decoder to reproduce the strain by inputting the latent vector and the label.

counterparts, i.e., about 4000 seconds for training and 1 milli second to generate a single waveform, thus for simplicity we will not listed here. Finally, one more point is that, after training is done, we will just choose the mean value of the latent layer to generate the waveform to avoid the stochastic feature of VAE.

TABLE VI. Summary of the accuracy for both LMR and HMR for the CVAE counterpart of each CAE waveform model considered in the main text.

	CVAE + NN	CVAE + CAE	2CVAE	1C2E1D	2C2E1D
Accuracy (LMR)	89.73%	89.03%	73.23%	94.35%	97.16%
Accuracy (HMR)	65.56%	70.26%	73.65%	79.11%	91.92%

[1] B. Abbott *et al.*, Observation of Gravitational Waves from a Binary Black Hole Merger, Phys. Rev. Lett. **116**, 061102 (2016).

[2] B. Abbott *et al.*, 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).

[3] R. Abbott *et al.*, GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run, 2020.

[4] B. J. Owen, Search templates for gravitational waves from inspiraling binaries: Choice of template spacing, Phys. Rev. D **53**, 6749 (1996).

[5] B. J. Owen and B. S. Sathyaprakash, Matched filtering of gravitational waves from inspiraling compact binaries: Computational cost and template placement, Phys. Rev. D **60**, 022002 (1999).

[6] B. F. Schutz, in *Data Processing, Analysis, and Storage for Interferometric Antennas* (Cambridge University Press, Cambridge, England, 1991), p. 406452.

[7] C. M. Biwer, C. D. Capano, S. De, M. Cabero, D. A. Brown, A. H. Nitz, and V. Raymond, Pycbc inference: A PYTHON-based parameter estimation toolkit for compact binary coalescence signals, Publ. Astron. Soc. Pac. **131**, 024503 (2019).

[8] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin *et al.*, Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library, Phys. Rev. D **91**, 042003 (2015).

[9] I. Hinder, The current status of binary black hole simulations in numerical relativity, Classical Quantum Gravity **27**, 114004 (2010).

[10] H. P. Pfeiffer, Numerical simulations of compact object binaries, Classical Quantum Gravity **29**, 124004 (2012).

[11] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer, Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms, Phys. Rev. Lett. **113**, 151101 (2014).

[12] S. Khan, K. Chatziioannou, M. Hannam, and F. Ohme, Phenomenological model for the gravitational-wave signal from precessing binary black holes with two-spin effects, Phys. Rev. D **100**, 024059 (2019).

[13] P. Ajith, M. Hannam, S. Husa, Y. Chen, B. Brgmann, N. Dorband, D. Mller, F. Ohme, D. Pollney, C. Reisswig, and *et al.*, Inspiral-Merger-Ringdown Waveforms for Black-Hole Binaries with Nonprecessing Spins, Phys. Rev. Lett. **106** (2011).

[14] A. Buonanno, Y. Pan, J. G. Baker, J. Centrella, B. J. Kelly, S. T. McWilliams, and J. R. van Meter, Approaching faithful templates for nonspinning binary black holes using the effective-one-body approach, Phys. Rev. D **76**, 104049 (2007).

[15] Y. Pan, A. Buonanno, M. Boyle, L. T. Buchman, L. E. Kidder, H. P. Pfeiffer, and M. A. Scheel, Inspiral-merger-ringdown multipolar waveforms of nonspinning black-hole binaries using the effective-one-body formalism, Phys. Rev. D **84**, 124052 (2011).

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

[17] A. Einstein, L. Infeld, and B. Hoffmann, The gravitational equations and the problem of motion, Ann. Math. **39**, 65 (1938).

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

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

[20] K. D. Kokkotas and B. G. Schmidt, Quasi-normal modes of stars and black holes, Living Rev. Relativity **2**, 2 (1999).

[21] H.-P. Nollert‘, Topical review: Quasinormal modes: The characteristic ‘sound’ of black holes and neutron stars, Classical Quantum Gravity **16**, R159 (1999).

[22] J. Centrella, J. G. Baker, B. J. Kelly, and J. R. van Meter, Black-hole binaries, gravitational waves, and numerical relativity, Rev. Mod. Phys. **82**, 3069 (2010).

[23] F. Loffler *et al.*, The Einstein toolkit: A community computational infrastructure for relativistic astrophysics, Classical Quantum Gravity **29**, 115001 (2012).

[24] J. Blackman, S. E. Field, C. R. Galley, B. Szilgyi, 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).

[25] M. Prrer, Frequency domain reduced order model of aligned-spin effective-one-body waveforms with generic mass ratios and spins, Phys. Rev. D **93**, 064041 (2016).

[26] D. Williams, I. Heng, J. Gair, J. 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).

[27] A. Nguyen, J. Clune, Y. Bengio, A. Dosovitskiy, and J. Yosinski, Plug & play generative networks: Conditional iterative generation of images in latent space, arXiv:1612.00005v2.

[28] K. Sohn, H. Lee, and X. Yan, Learning structured output representation using deep conditional generative models, in *Advances in Neural Information Processing Systems*, (Curran Associates, Inc., 2015), https://papers.nips.cc/paper/2015/hash/8d55a249e6baa5c06772297520da2051-Abstract.html.

[29] F. Tonolini, J. Radford, A. Turpin, D. Faccio, and R. Murray-Smith, Variational inference for computational imaging inverse problems, arXiv:1904.06264v3.

[30] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith, Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy, arXiv:1909.06296.

[31] S. R. Green and J. Gair, Complete parameter inference for GW150914 using deep learning, arXiv:2008.03312.

[32] S. R. Green, C. Simpson, and J. Gair, Gravitational-wave parameter estimation with autoregressive neural network flows, Phys. Rev. D **102**, 104057 (2020).

[33] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborov, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[34] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, USA, 2016), http://www.deeplearningbook.org.

[35] G. E. Hinton and R. S. Zemel, Autoencoders, minimum description length and Helmholtz free energy, in *Proceedings of the 6th International Conference on Neural Information Processing Systems, NIPS'93, San Francisco, CA, USA, 1993* (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993), pp. 3–10, https://dl.acm.org/doi/proceedings/10.5555/2987189.

[36] D. P. Kingma and M. Welling, An introduction to variational autoencoders, arXiv:1906.02691.

[37] R. Yu, A tutorial on VAEs: From Bayes' rule to lossless compression, arXiv:2006.10273v2.

[38] F. Chollet, *Deep Learning with* PYTHON (Manning Publications Co., New York, 2017).

[39] S. Babak, H. Fang, J. R. Gair, K. Glampedakis, and S. A. Hughes, 'Kludge' gravitational waveforms for a test-body orbiting a Kerr black hole, Phys. Rev. D **75**, 024005 (2007); Erratum, Phys. Rev. D **77**, 049902 (2008).

[40] J. Lee, S. H. Oh, K. Kim, G. Cho, J. J. Oh, E. J. Son, and H. M. Lee, Deep learning model on gravitational waveforms in merging and ringdown phases of binary black hole coalescences, arXiv:2101.05685v2.