PROCEEDINGS OF SCIENCE

Comparison of training methods for Convolutional Neural Network model for Gravitational-Wave detection from Neutron Star−Black Hole Binaries

Suyog Garg,[a,b,*] **Seiya Sasaoka,**[c] **Diego Dominguez,**[c] **Yilun Hou,**[c] **Naoki Koyama,**[d] **Kentaro Somiya,**[c] **Hirotaka Takahashi**[b,e] **and Masatake Ohashi**[a,b]

[a] *Department of Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan 113-0033*
[b] *Institute of Cosmic Ray Research, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa City, Chiba, Japan 277-8582*
[c] *Department of Physics, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro-ku, Tokyo, Japan 152-8551*
[d] *Graduate School of Science and Technology, Niigata University, 8050 Ikarashi-2-no-cho, Nishi-ku, Niigata City, Niigata, Japan 950-2181*
[e] *Research Center for Space Science, Advanced Research Laboratories, Tokyo City University, 3-3-1 Ushikubo-nishi, Tsuzuki-ku, Yokohama, Kanagawa, Japan 224-8551*

E-mail: gargsuyog@g.ecc.u-tokyo.ac.jp, sasaoka@gw.phys.titech.ac.jp, diego@gw.phys.titech.ac.jp, elon@gw.phys.titech.ac.jp, f221004a@mail.cc.niigata-u.ac.jp, somiya@phys.titech.ac.jp, hirotaka@tcu.ac.jp, ohashi@icrr.u-tokyo.ac.jp

The traditional method of Gravitational Wave (GW) detection is Matched Filtering that was used for the first GW detection by aLIGO in 2015. The method works by matching the observation data sample with a set of templates of known GW waveforms. Iterating through all the templates for relatively complex GW signals, for instance those from eccentric sources, increases the overall computational cost and time complexity. In recent years, Machine Learning techniques have been probed as a solution to this problem. In this short paper, we present a new Convolutional Neural Network model for detection of GW signals from Neutron Star−Black Hole (NSBH) binaries in Gaussian random noise. We use NSBH signals simulated using IMRPhenomNSBH LALsuite waveform approximant for training the model. We then compare the model detection sensitivities for three different training strategies obtained by combining Uniform and Non-uniform Signal-to-Noise distribution in the training dataset with the Curriculum Learning training methodology. We find that two of the training strategies perform considerably better than the other one for all test datasets considered.

38th International Cosmic Ray Conference (ICRC2023)
26 July - 3 August, 2023
Nagoya, Japan

ICRC2023
38th International Cosmic Ray Conference
The Astroparticle Physics Conference

*Speaker

PoS(ICRC2023)1536

1. Introduction

Gravitational Waves (GW) are ripples in the fabric of spacetime caused by spherically non-symmetric motion of a body of mass. The GW signals are extremely weak and just the ones created by astrophysical processes are strong enough to be detected on Earth. They are also the only possible way of observing the very early Universe, since it is opaque to electromagnetic radiation. Gravitational Waves are a robust prediction of the General Theory of Relativity formulated by Einstein in 1915. However, it was only the development of large-scale laser interferometry based GW detectors in the last decades of the 20th century, that paved the way for direct observations of Gravitational Waves. Three such ground based laser interferometric GW observatories are currently in operation: advanced LIGO (Hanford and Livingston), advanced Virgo and KAGRA. Advanced LIGO performed the first direct detection of Gravitational Waves by the 2015 detection of the GW150914 signal originating from the merger of two Black Holes [1]. This phenomenal detection ushered in the era of Gravitational Wave Astronomy. Since then around 100 such signals have been detected by the advanced LIGO and the advanced Virgo.

A well-established method for Gravitational Wave detection is the matched filtering technique that was used for the GW150914 signal detection. The matched filtering technique involves convolving the detector data with samples from a pre-calculated template bank containing a set of expected Gravitational Wave signal waveforms. The Gravitational Wave templates are mathematically generated by altering parameter values describing the source of Gravitational Waves, for instance, their mass, spin and orientation. Traditional matched filtering based algorithms are the most sensitive methods for Gravitational Wave detection. However, they require a large execution time as the size of the template bank increases. This is especially true when the GW signals contain effects like higher-order modes, precession and orbital eccentricity at the source.

To counter these shortcomings of the matched filtering method, Machine Learning based techniques for GW detection have been gaining momentum in recent years. Machine Learning is a field of Artificial Intelligence devoted to developing models that “learn” to efficiently perform any particular task. The “learning” happens by training the model on a set of training data, so that it can make predictions on another similar set of test data. The models train to make accurate predictions by optimizing a model critical function through a supervised or unsupervised learning approach.

The first application of Machine Learning in GW science was done by George and Huerta [3]. They demonstrated that Machine Learning can really be useful for GW detection and parameter estimation. Since then several studies have further been done to assess the feasibility of using Machine Learning techniques for detection of Gravitational Waves. In this short article, we present a new Convolutional Neural Network (CNN) for detection of GWs from Neutron Star−Black Hole binary sources. We also make a comparison between different training strategies used for training the model. In Section 2.1, we lay out our CNN model architecture. Next in Section 2.2, we describe how the training and test data for the model is generated and processed. Then in explain the Section 2.3, the Curriculum Learning training methodology and three different training strategies that are used for model training. Finally, in Section 3 we present our results from the model testing on test datasets and discuss possible future research directions.

PoS(ICRC2023)1536

2. Model Training

2.1 Network Architecture

We use a simple Residual network architecture for our Convolutional Neural Network model for detection of Gravitational Waves from Neutron Star–Black Hole binary sources. The network consists of four Residual blocks, each of which contain one Convolutional layer and one Max Pooling layer. The Residual blocks are followed by a flattening layer, which are then followed by three dense layers. The input to the network is a GW time series (the input dimension 2 corresponds to the 2 detectors used to generate the training data). The output of the network is an array of length two, corresponding to the label: Signal or No-Signal. The network architecture of our model is outlined in Figure 1. Additionally, Table 1 tabulates the output shape and the number of parameters in each layer of the network.

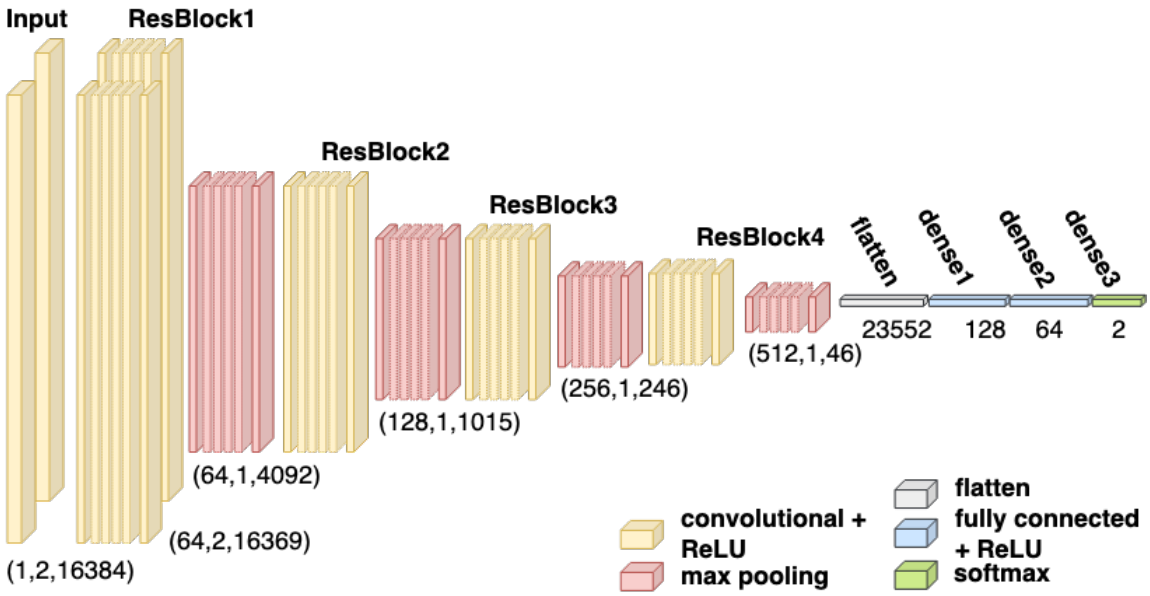

Figure 1: Network architecture of the Convolutional Neural Network model Gravitational Wave detection from Neutron Star–Black Hole binary sources. The network consists of four Residual blocks, each containing one Convolutional and one Max Pooling layer. The output shape at each block is also shown. The input to the network is a Gravitational Wave time series on the left, while the output of the network is the label (signal or no-signal) on the right.

2.2 Training Data

To create the training samples for the Neutron Star–Black Hole model (Figure 1), we use the IMRPhenomNSBH waveform approximant available via PyCBC. The IMRPhenomNSBH approximant includes tidal (on the Neutron Star component of the binary), spin and precession effects into the modelling of the waveform. Incorporating tidal effects within the equations describing the evolution of the system is important for Neutron Star–Black Hole mergers, especially for large mass ratio binaries [4]. Using this approximant, we generate Neutron Star–Black Hole GW signal waveforms the two advanced LIGO detectors, H1 and L1.

Table 1: Model parameters for the Convolutional Neural Network network used for Gravitational Wave detection from Neutron Star–Black Hole binary sources. The network consists of four Residual blocks, each containing one Convolutional and one Max Pooling layer. The full network architecture is shown in Figure 1.

Layer (type)	Output Shape	Param #
batch_normalization (BatchNormalization)	(None, 1, 2, 16384)	65536
conv2d (Conv2D)	(None, 64, 2, 16369)	1088
max_pooling2d (MaxPooling2D)	(None, 64, 1, 4092)	0
conv2d_1 (Conv2D)	(None, 128, 1, 4062)	131200
max_pooling2d_1 (MaxPooling2D)	(None, 128, 1, 1015)	0
conv2d_2 (Conv2D)	(None, 256, 1, 985)	524544
max_pooling2d_2 (MaxPooling2D)	(None, 256, 1, 246)	0
conv2d_3 (Conv2D)	(None, 512, 1, 184)	4194816
max_pooling2d_3 (MaxPooling2D)	(None, 512, 1, 46)	0
flatten (Flatten)	(None, 23552)	0
dense (Dense)	(None, 128)	3014784
dense_1 (Dense)	(None, 64)	8256
dense_2 (Dense)	(None, 2)	130
Total params 7,940,354		
Trainable params 7,907,586		
Non-trainable params 32,768		

The signals are generated for a total duration of 4.25 sec and are sampled at 4096 Hz. Signals of longer duration are required because the inspiral time for Neutron Star–Black Hole is much longer than that of Binary Black Hole, since the binary component masses have a larger difference. Then to create training samples with signals embedded in noise, we add Gaussian random noise to these generated signal waveforms. Based on the training requirements, some pure Gaussian noise samples are also generated. All these generated samples are then whitened using the `aLIGOZeroDetHighPower` PSD to match the actual detector power spectrum. After whitening the sample duration gets reduced to 4.0 sec. The left panels on Figure 2 illustrates some representative training samples used in our analysis.

For our model we use two different Signal-to-Noise (SNR) distributions in the training data, namely: Uniform and Non-Uniform SNR distributions. For each of our SNR distributions in the training data, we generate the waveforms for a Black Hole mass range of $m_1 = 5 - 20M_\odot$ and a Neutron Star mass range of $m_2 = 1 - 2M_\odot$. The Uniform SNR distribution is the Gaussian random process, where the probability of occurrence of different variables is the same. This is shown on the top right panel in Figure 2. George and Huerta [3] also use the Uniform SNR distribution for training their model.

LIGO observations till date have been around SNR ~ 6. So, having more number of training samples with SNR in this neighbourhood will be preferable. That's idea behind using a Non-uniform SNR distribution. Qiu et al. [5] use a truncated triangular SNR distribution with left=3, mode=5 and right=27.5. Whereas, George and Huerta [3] used Curriculum learning technique by training with large SNR samples first, while the samples were uniformly sampled from a uniform SNR

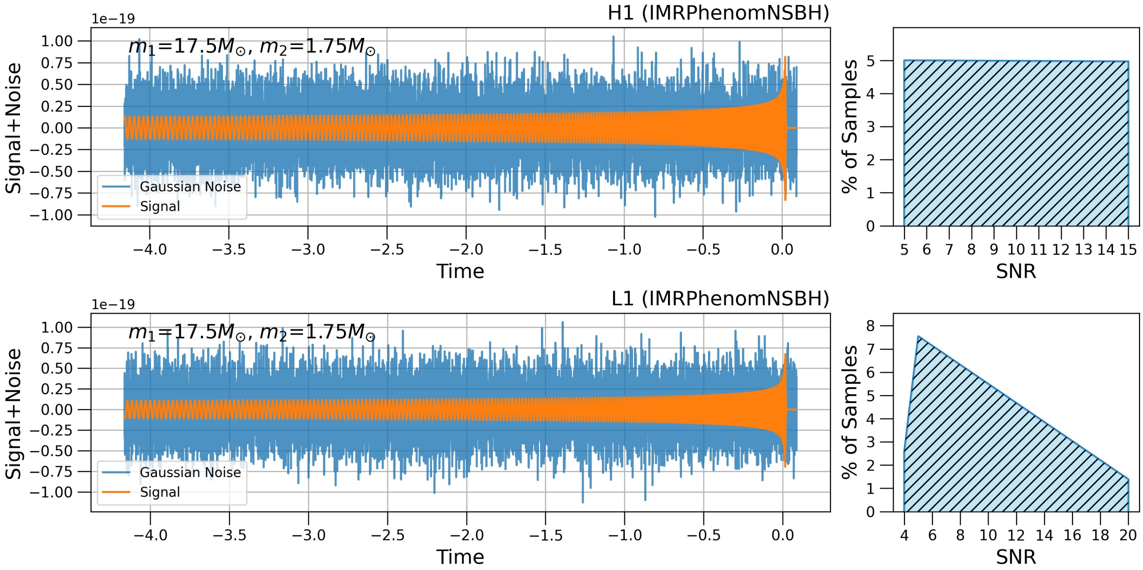

Figure 2: *Left:* A representative Neutron Star–Black Hole Gravitational Wave signal waveform used for training our model. The waveforms are generated using the IMRPhenomNSBH waveform approximant available via PyCBC. *Right:* Two Signal-to-Noise (SNR) ratio distributions in the training data used for training our model. The distribution in the Top has the number of samples uniformly sampled for all SNR and corresponds to the "Uniform SNR distribution". The distribution in the bottom plot is the "Non-uniform SNR distribution" taken from [5].

distribution of the training data. We use the Qiu et al. [5] SNR distribution as our Non-uniform SNR distribution. This distribution is shown on the bottom right panel in Figure 2.

2.3 Curriculum Learning and Training Strategies

There can be a number of different methods to train a Convolutional Neural Network model. Curriculum Learning is a well-known training methodology used for Convolutional Neural Networks to improve their model accuracy during training with complex datasets. The Curriculum Learning training methodology was first introduced by Bengio et al. [2] in their, "Discussing merits and demerits of various training methodologies for Artificial Neural Networks". In the field of GW science, Curriculum Learning has been used for Machine Learning applications to Binary Black Hole GW detections. The methodology has been found to increase the sensitivity (and accuracy) of the model for data samples containing a GW signal at low SNR [3].

The key idea of Curriculum Learning is to train a Convolutional Neural Network model by gradually increasing in the complexity of the training data. For the Gravitational Wave data, a simple data sample corresponds to the sample with a high SNR, while a more complex data is the one with the signal at low SNR. Thus, when using the Curriculum Learning training methodology, we first train the model with high SNR samples and follow this by using samples at lower SNRs in gradual steps. The full Curriculum Learning algorithm used for our model is laid out in Algorithm 1 below.

Algorithm 1 The Curriculum Learning Algorithm

Generate pure signals strains for $m_{\mathrm{BH}} \in (5, 20)\ M_\odot$ and $m_{\mathrm{NS}} \in (1, 2)\ M_\odot$
for $k \in [20.0, 16.0, 13.5, 12.0, 9.0, 7.5, 6.0, 4.0, 3.0, 2.0, 1.5, 1.0]$ **do**
  if `distro` = `uniformSNRdistro` **then**
    `snr` ← `random(uniformSNRdistro(5,15))`
  else
    `snr` ← `random(quiSNRdistro())`
  end if
  `target_strain` ← `strain` × (`snr` × k) / `network_snr`
  `sample` ← `noise` + `target_strain` ▷ whiten sample
end for
Test trained model with test dataset having `snr` $\in (17, 2, -0.5)$

For our model we use three different training strategies obtained by combining the Uniform and Non-uniform SNR distributions with the Curriculum Learning training methodology. These strategies are: Uniform SNR distribution with Curriculum Learning, Non-uniform SNR distribution without Curriculum Learning and Non-uniform distribution with Curriculum Learning.

3. Results

We use identical training datasets for each of our training strategies. The datasets differ only in the randomly generated noise profiles. The number of training samples is fixed at 50,000 samples, out which 25% samples only contain pure noise. The remaining samples have GW signal waveforms embedded in the Gaussian random noise. We train the model for a total of 10 training epochs with a batch size of 50. For the learning rate α, we follow a schedule of $\alpha = 10^{-3}$ for epochs $1-5$ and $\alpha = \gamma \times 10^{-3}$ for the remaining epochs, where $\gamma = 0.1$ for epochs $5-7$ and $\gamma = 0.01$ for epochs $7-10$.

Once the model has finished training, we test the model on test datasets for signals at different SNRs. We generate the test datasets with test SNRs in the range 1 to 17 with a step-size of 0.5. The results for testing the model training for each of our training strategies are presented in Figure 3, in the form of a SNR versus Sensitivity (and Accuracy) curves. To calculate the Sensitivity and Accuracy we use a fixed False Alarm Rate value (FAR) of FAR = 0.06, which is a reasonable threshold for the detections.

Figure 3 plots the model Sensitivity (left panel) and Accuracy (right panel) versus SNR curves for the three training strategies considered in this study. It is clearly seen that both the training strategies with Curriculum Learning perform equally well for the Neutron Star–Black Hole GW detection in Gaussian noise. However, when Curriculum Learning is not used we find a substantial decline in the Sensitivity and Accuracy of the model. The Uniform SNR distribution with Curriculum Learning model obtains more than 90% Sensitivity and Accuracy at SNR >= 12. The Sensitivity attains near saturation at SNR > 14. On the other hand, the Non-uniform SNR distribution without Curriculum Learning model performs significantly worse as compared to the Uniform SNR distribution with Curriculum Learning case. This model obtains more than 90%

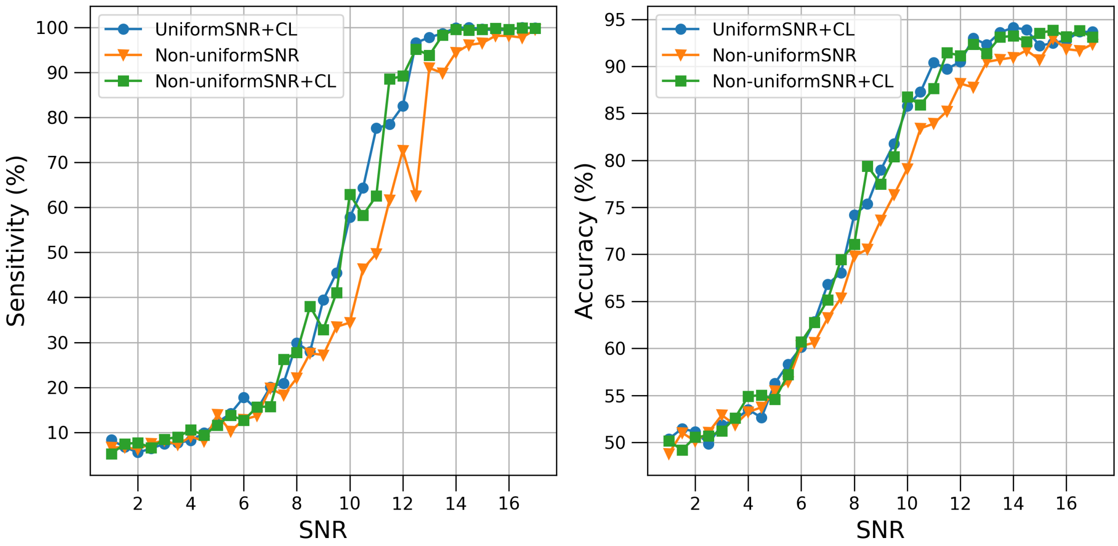

Figure 3: Comparison of model Sensitivity (left) and Accuracy (right) at different SNR values for the three training strategies. Curriculum Learning is seen to perform the best regardless of the SNR distribution of the data.

Sensitivity and Accuracy at SNR >= 12, while the Sensitivity attains saturation at all SNR > 14. The performance of the Non-uniform SNR distribution with Curriculum Learning model is similar to the Uniform SNR distribution with Curriculum Learning model, with more than 90% Sensitivity and Accuracy at SNR >= 12 and saturation at SNR > 14.

In the fourth LIGO-Virgo-KAGRA observation run that commenced in May 2023, the best expected SNR for GW detections is at SNR = 8.[1] Around this SNR, both the training strategies that follow Curriculum Learning training methodology perform better than the one without it, even though the Sensitivity is only around 30%. If the FAR constraint is relaxed, say to have FAR = 1, better Sensitivity and Accuracy at low SNR will be achieved, but at the expense of more number of false alarms. Further, the Curriculum Learning strategies reache close to saturation, that is all $\geq$ 90% signals are detected, after SNR = 12, whereas the training strategy without Curriculum Learning reaches similar Sensitivity at SNR = 14 and beyond. It also be noted that Curriculum Learning with Non-Uniform SNR distribution requires almost 5 times more running time as compared to the other training strategies.

Our main result is that Curriculum Learning improves the model Sensitivity regardless of the SNR distribution in the training data. This has clear implications for training future Machine Learning based methods for detection of GW signals in noisy data. Our future work will focus on detection GW signals embedded in real detector noise. We shall continue to use the Curriculum Learning methodology for training the model. Real noise data from O3 observation run can be obtained from the Gravitational Wave Open-Science Center `https://www.gw-openscience.org/O3/O3b/` data files by selecting the time range in which no events were observed.

[1]See: https://emfollow.docs.ligo.org/userguide/capabilities.html

Acknowledgments

This research was supported in part by JSPS Grant-in-Aid for Scientific Research [No. 22H01228 (K. Somiya), and Nos. 19H01901, 23H01176 and 23H04520 (H. Takahashi)]. This research was also supported by the Joint Research Program of the Institute for Cosmic Ray Research, University of Tokyo and Tokyo City University Prioritized Studies. This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gwosc.org), a service of the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. This material is based upon work supported by NSF's LIGO Laboratory which is a major facility fully funded by the National Science Foundation, as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded, through the European Gravitational Observatory (EGO), by the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN) and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. KAGRA is supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan.

References

[1] B. P. Abbott, R. Abbott, T. D. Abbott, et al. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 116:061102, Feb 2016. URL https://link.aps.org/doi/10.1103/PhysRevLett.116.061102.

[2] Y. Bengio, J. Louradour, R. Collobert, and J. Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, jun 2009. URL https://doi.org/10.1145%2F1553374.1553380.

[3] D. George and E. A. Huerta. Deep neural networks to enable real-time multimessenger astrophysics. *Phys. Rev. D*, 97:044039, Feb 2018.

[4] J. E. Thompson, E. Fauchon-Jones, S. Khan, et al. Modeling the gravitational wave signature of neutron star black hole coalescences: Phenomnsbh. *LIGO-P2000059*, 02 2020. URL https://arxiv.org/pdf/2002.08383.pdf.

[5] R. Qiu, P. Krastev, K. Gill, and E. Berger. Deep Learning Detection and Classification of Gravitational Waves from Neutron Star-Black Hole Mergers. *arXiv e-prints*, page arXiv:2210.15888, Oct. 2022, 2210.15888.

PoS(ICRC2023)1536