


Efficient parameter inference for gravitational wave signals in the presence of transient noises using temporal and time-spectral fusion normalizing flow*

Tian-Yang Sun (孙天阳)¹ Chun-Yu Xiong (熊春雨)¹ Shang-Jie Jin (金上捷)¹ Yu-Xin Wang (王钰鑫)¹
Jing-Fei Zhang (张敬飞)¹ Xin Zhang (张鑫)^{1,2,3†} 

¹Key Laboratory of Cosmology and Astrophysics (Liaoning) & College of Sciences, Northeastern University, Shenyang 110819, China

²Key Laboratory of Data Analytics and Optimization for Smart Industry (Ministry of Education), Northeastern University, Shenyang 110819, China

³National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University, Shenyang 110819, China

Abstract: Glitches represent a category of non-Gaussian and transient noise that frequently intersects with gravitational wave (GW) signals, thereby exerting a notable impact on the processing of GW data. The inference of GW parameters, crucial for GW astronomy research, is particularly susceptible to such interference. In this study, we pioneer the utilization of a temporal and time-spectral fusion normalizing flow for likelihood-free inference of GW parameters, seamlessly integrating the high temporal resolution of the time domain with the frequency separation characteristics of both time and frequency domains. Remarkably, our findings indicate that the accuracy of this inference method is comparable to that of traditional non-glitch sampling techniques. Furthermore, our approach exhibits a greater efficiency, boasting processing times on the order of milliseconds. In conclusion, the application of a normalizing flow emerges as pivotal in handling GW signals affected by transient noises, offering a promising avenue for enhancing the field of GW astronomy research.

Keywords: gravitational wave, glitch, non-Gaussian and transient noise, normalizing flow, machine learning, likelihood-free parameter inference

DOI: 10.1088/1674-1137/ad2a5f

I. INTRODUCTION

Since the initial detection of gravitational waves (GWs) stemming from binary black hole (BBH) mergers [1] and a binary neutron star (BNS) event [2–4], the LIGO-Virgo-KAGRA collaboration [5–7] has reported the detections of over 90 GW events involving compact binary coalescences. These GW observations play a pivotal role in advancing fundamental physics [8–13], astrophysics [14–19], and cosmology [20–37]. However, it is important to note that GW data are susceptible to non-Gaussian noise contamination [38–41], which potentially influences the aforementioned scientific analyses.

Non-Gaussian noise, referred to as glitches and non-stationary noise, poses a challenge due to it contaminating GW data, thereby impacting the quest for GW signals and the accurate estimation of wave source parameters [41–46]. Various sources contribute to glitches, including seismic events such as earthquakes [47] and environmental factors like passing trains [48]. While nu-

merous techniques are presently employed to mitigate glitches at their source [49–52], future observations operating at heightened sensitivities are anticipated to yield increased detection rates [53–55]. Consequently, the reduction of GW signals tainted by glitches becomes a challenging task [56].

During the recent third LIGO-Virgo observing run (O3), 24% of the GW candidates were found to be contaminated by glitches [39, 40]. Notably, the analyses of prominent events like GW170817, GW200129, and GW191109_010717 were substantially impacted due to glitches [11, 57–59]. Therefore, effective glitch mitigation strategies are imperative prior to undertaking parameter estimation.

The two primary methods commonly employed for complete glitch removal are BayesWave [60, 61] and gwsbtract [62, 63]. Additionally, various approaches exist for mitigating the effects of glitches in observational data [46, 64–70]. However, as of O3, there remains a lack of reasonably low-latency methods for data cleaning [42],

Received 19 January 2024; Accepted 19 February 2024; Published online 20 February 2024

* Supported by the National SKA Program of China (2022SKA0110200, 2022SKA0110203), the National Natural Science Foundation of China (11975072, 11875102, 11835009), and the National 111 Project (B16009)

† E-mail: zhangxin@mail.neu.edu.cn (Corresponding author)

©2024 Chinese Physical Society and the Institute of High Energy Physics of the Chinese Academy of Sciences and the Institute of Modern Physics of the Chinese Academy of Sciences and IOP Publishing Ltd

posing a potential hindrance to the discovery of certain physical phenomena, such as the observation of subsequent electromagnetic (EM) counterparts following BNS mergers. This challenge arises from the computational intensity of the full Bayesian approach, making it a time-consuming process. Consequently, there is a pressing need to develop precise and low-latency deglitching methods.

One prevalent technique for alleviating the impact of corrupted glitches is gating [71], enabling the prompt removal of glitch-affected data with low latency [72]. A notable case is the handling of GW170817, for which contaminated segments of the data were excised to facilitate the search for EM counterparts [57]. However, the confluence of gating and signals in the time-frequency domain poses challenges to parameter inference [57]. Machine learning emerges as a promising solution for addressing this issue, given its non-linear GPU-based computational capabilities, making it well-suited for the low-latency processing of non-stationary data [73]. In addition, the robustness of machine learning also makes it more suitable for processing GW data contaminated by glitches [74–77].

Several studies have explored the application of machine learning in reconstructing glitches from data, enabling subsequent subtraction to mitigate their effects [78–84]. Following the acquisition of clean data, additional computations are often necessary for deriving source parameters for subsequent analyses [85]. Consequently, we aim to address the query of whether machine learning can be effectively employed to directly infer GW parameters from contaminated data.

Neural posterior estimation [86–89], relying on a normalizing flow, demonstrates a precise estimation of the posterior distribution of the source parameters [90–92]. Functioning as a likelihood-free method, a normalizing flow proves effective in handling non-Gaussian data, exemplified by its successful application to the 21 cm signal [93–96]. Consequently, our exploration aims to ascertain the viability of employing a normalizing flow in the processing of GW data contaminated by glitches.

In this study, we introduce a novel method grounded in a normalizing flow for parameter estimation for data afflicted by glitches. While utilizing time-frequency domain data proves advantageous for glitch-contaminated data [97–99], inherent limitations in the time-frequency resolution and binning may result in the loss of intricate details, thereby influencing the parameter inference. In recent years, some researchers have studied improvements of the network performance from the data fusion of time series and corresponding frequency domain data [100–102]. Therefore we employ a dual approach, incorporating both time-domain and time-frequency domain data in the parameter inference process, i.e., temporal and time-spectral fusion normalizing flow (TTSF-NF). Our

investigation specifically targets high signal-to-noise ratio (SNR) glitches that defy resolution through robustness.

The organization of this work is as follows: Section II provides an introduction to the methodology employed. In Section III, we comprehensively present the results yielded by our approach. The conclusion is encapsulated in Section IV.

II. METHODOLOGY

A. Data generation

In this study, we focused on two prevalent glitches commonly observed in aLIGO detectors — namely, “blip” and “scattered light”— both sourced from Gravity Spy [103]. The spectrogram diagrams for each glitch are illustrated in Fig. 1. Our glitch selection specifically targeted segments with an SNR exceeding 12, a criterion chosen due to the prevalent focus of existing research on scenarios with a relatively low SNR [46, 104]. As depicted in Fig. 2, relying solely on the robustness of the time normalizing flow (T-NF) proves insufficient for resolving this challenge. Notably, in the case of Hanford’s blip during O3b, the proportion of data with $\text{SNR} > 12$ is 54%. This underscores the likelihood of encountering simultaneous occurrences of a signal and a high SNR glitch in future observations.

Blip glitches exhibit a brief duration (~ 0.1 ms) and a frequency ranging from tens of Hz to hundreds of Hz, precisely aligning with the frequency range of BBH [105]. Their shape easily resembles that of GW signals [105, 106]. The origin of blip glitches is presently unknown, and they do not manifest in any auxiliary channels [107]. Consequently, it is probable that they will continue to be a significant component of glitches in the future.

Scattered light glitches typically exhibit persistence and a low frequency. This type of noise arises when a fraction of the laser beam light scatters due to defects in the detector mirror or excessive ground motion, subsequently recombining with the main beam [48, 108]. Although its occurrence diminishes with advancements in the manufacturing processes [40], it tends to escalate with increased sensitivity. Consequently, achieving complete elimination in the future is likely to be challenging.

In our approach, we utilized data with a duration of 2 s. Considering the higher accuracy in determining the merger time during signal search [109, 110], we held the GPS time of the merger relative to the center of the Earth fixed at $t_0 = 1.8$ s and then projected it onto the detector based on the corresponding parameters. Leveraging the work of Alvarez-Lopez *et al.* [111], who introduced a model capable of concurrently identifying glitches and GW signals in the data, we conducted separate training

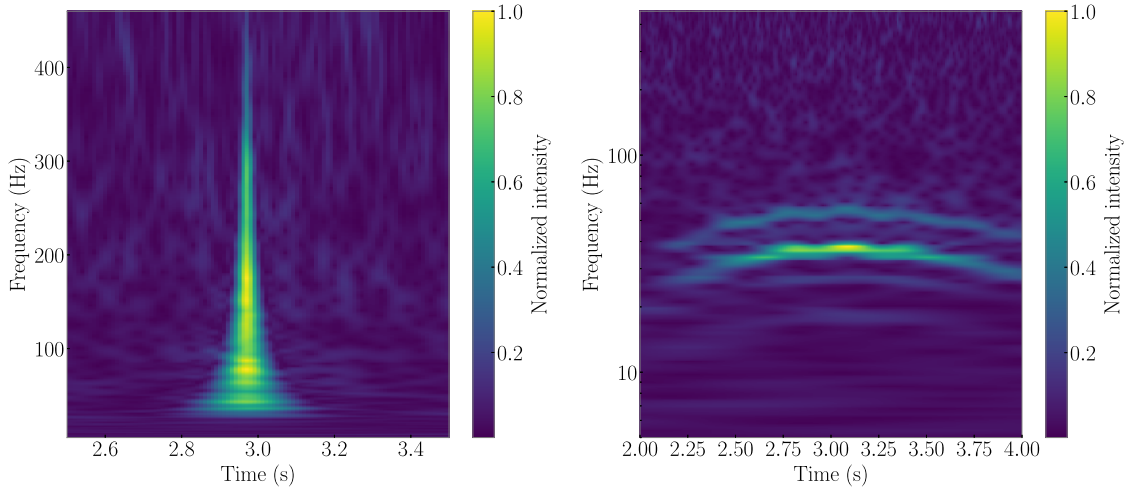


Fig. 1. (color online) Two types of glitches considered in this work. Left panel: Blip noise characterized by durations on the order of milliseconds and a broad frequency bandwidth on the order of 100 Hz. Right panel: scattered light noise which persists for an extended duration and exhibits a frequency below 100 Hz.

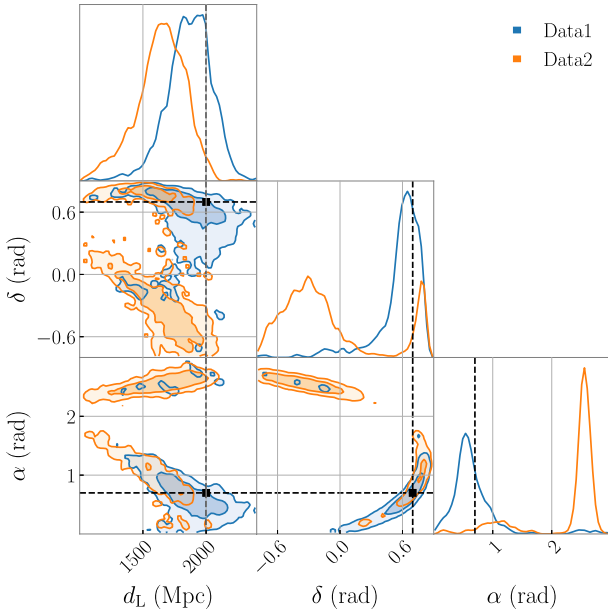


Fig. 2. (color online) One- and two-dimensional marginalized posterior distributions for d_L , δ , and α using Data1 and Data2. The intersection points of the dashed lines are the injected parameters. Note that Data1 are the posterior parameters predicted by T-NF with Gaussian noise plus the GW signal and Data2 are the posterior parameters predicted by T-NF with Gaussian noise with blip plus the GW signal.

for each glitch type. To replicate the coincidence of the glitch and the merger moment, we injected glitches at a temporal distance of $(-0.1 \text{ s}, +0.1 \text{ s})$ from the merger moment. The time strain data can be written as

$$s(t) = h(t) + n(t), \quad (1)$$

where $h(t)$ is the GW signal and $n(t)$ is the background noise (including the glitch). The specific time strain is shown in Fig. 3.

Referring to Hourihane *et al.* [68], glitches have a more pronounced impact on high-quality systems compared with on low-quality systems. Consequently, this study primarily focuses on higher mass BBH systems. Several representative parameters were considered, and the expectation is that our TT-SF-NF will demonstrate applicability to other parameter sets as well. Specific parameters are detailed in Table 1, where $f_{\max} < 512 \text{ Hz}$, and events with $8 < \text{SNR} < 32$ were exclusively selected. Given the substantial amount of existing observational data, we employed real Hanford noise containing glitches, calibrated according to Gravity Spy. The noise in other detectors corresponds to the glitch-free segments obtained from Gravitational Wave Open Science Center (GWO-SC) [112]. Adhering to the Nyquist sampling law, the data's frequency only needs to exceed 2 times the value of f_{\max} ; hence, we downsampled the data to 1024 Hz. We introduced the GW signal into the noise, subsequently whitening it to derive the simulated time-domain data.

The spectrogram utilized in this study was generated through the Q -transform [113], a modification of the short Fourier transform. It maps the time-frequency plane with pixels of constant Q (quality factor) by employing an analysis window whose duration is inversely proportional to the frequency. Recognizing the necessity for real-time data analysis, the spectrogram's resolution is constrained to maintain efficiency. To accommodate TT-SF-NF's requirements, time and frequency were divided into 200 bins to achieve the necessary image size. The observational data d input to TT-SF-NF consists of both time strain data and the spectrogram. All these steps were implemented using PyCBC [114], with the BBH wave-

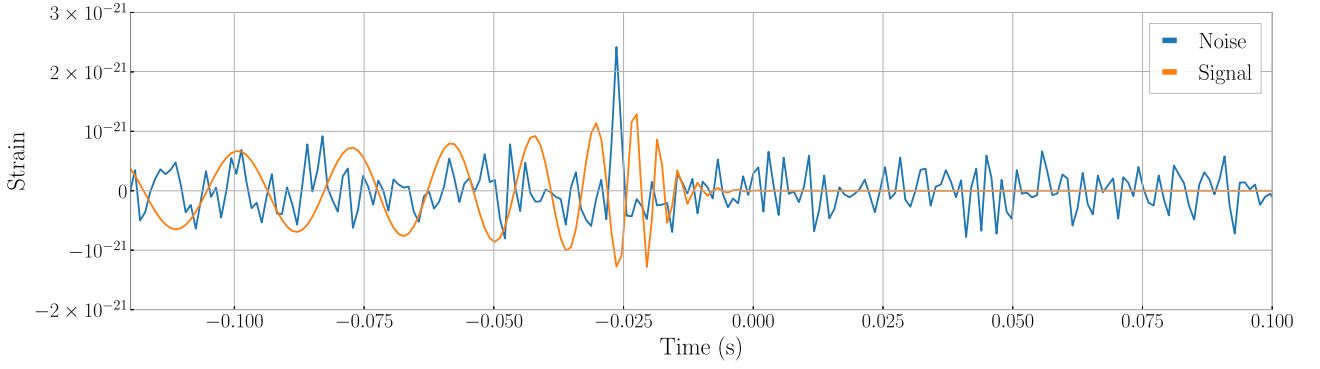


Fig. 3. (color online) Time-domain strains of noise and GW signal. Here, the noise strain data is adopted from LIGO Hanford noise and the GW signal is generated based on the GW150914-like event. Note that the blip is injected at $t_0 - 30$ ms of the GW signal (this is a situation depicted in Figure 3 of Ref. [42]).

Table 1. Distribution of simulated GW waveform parameters. Note that other parameters not mentioned are set to zero for simplicity.

Parameter	Uniform distribution
Chirp mass	$\mathcal{M}_c \in [25.0, 62.5]M_\odot$
Mass ratio	$q \in [0.5, 1]$
Right ascension	$\alpha \in [0, 2\pi]$ rad
Declination	$\delta \in [-\pi/2, \pi/2]$ rad
Polarization angle	$\psi \in [0, 2\pi]$ rad
Luminosity distance	$d_L \in [300, 3000]$ Mpc

form modeled using IMRPhenomPv2 [115].

B. Normalizing flow

The backbone of TTFSF-NF is based on the normalizing flow (NF) [86], a generative model. The fundamental concept of NF involves providing a reversible transformation f_d for each observation data d , thereby converting the simple base distribution into a more intricate posterior distribution. The pivotal aspect of this network lies in the reversibility of the transformation and the straightforward computation of the Jacobian matrix. Currently, NF has found extensive application in GW signal processing [90, 104, 116–120].

The NF can be expressed by the following formula [86]

$$q(\theta | d) = \pi(f_d^{-1}(\theta)) |\det J_{f_d^{-1}}|. \quad (2)$$

The basic distribution $\pi(u)$ can, in principle, be arbitrary. However, for ease of sampling and density evaluation, it is often chosen to be a standard multivariate normal distribution with the same dimension D as the sample space. In practice, Eq. (2) can be continuously applied to construct arbitrarily complex densities [86]

$$f_d(u) = f_{d,N} \circ f_{d,N-1} \circ \cdots \circ f_{d,1}(u), \quad (3)$$

where $f_{d,i}(u)$ ($i = 1, \dots, N$) represents a block of the NF. This approach is referred to as neural posterior estimation (NPE). The objective of NPE is to train a parameter conditional distribution that approximates the true posterior distribution. This task translates into an optimization problem with the aim of minimizing the expected Kullback-Leibler (KL) divergence [121] between these two distributions.

The loss of NF can be written as the expected value (over d) of the cross entropy between the true and model distributions [122]

$$L = \mathbb{E}_{p(d)} [\text{KL}(p(\theta | d) || q(\theta | d))]. \quad (4)$$

On a minibatch of training data of size N , we approximate [116]

$$L \approx -\frac{1}{N} \sum_{i=1}^N \log q(\theta^{(i)} | d^{(i)}). \quad (5)$$

It is evident that the training of NF necessitates only the parameters corresponding to each data point without making any assumptions. This constitutes a likelihood-free method, eliminating the need for data modeling. Hence, it is well-suited for non-Gaussian and other challenging model scenarios.

In this study, we employed a currently more potent flow known as the neural spline flow (NSF) [123]. The fundamental concept behind NSF revolves around a fully differentiable module founded on monotonic rational-quadratic splines. Specifically, it consists of a series of coupling transforms. For n -dimensional data, the coupling transform maps the input vector x (for the $(i+1)$ th block, $x = f_{d,i} \circ \cdots \circ f_{d,1}(u)$) to the output y in the following way [116, 120].

1. Divide the input x into two parts, $x = [x_{1:m}, x_{m+1:n}]$ where $m < n$.

2. Input $x_{1:m}$ and the results output by the preceding neural network based on the input data into another neural network (specifically, the residual neural network) to obtain the vector α .

3. For each α_i ($i = m, \dots, n$), one can construct the invertible function g_{α_i} for computing $y_i = g_{\alpha_i}(x_i)$.

4. Set $y_{1:m} = x_{1:m}$.

Finally, return $y = [y_{1:m}, y_{m+1:n}]$. The neural spline coupling transform, meanwhile, treats each output as a monotonically increasing segmented function, and by putting the abovementioned α_i through a series of operations one can obtain the knots $\{(u_i^{(k)}, c_i^{(k)})\}_{k=0}^K$ and the positive-valued derivatives $\{\delta_i^{(k)}\}_{k=0}^K$, which can be expressed in terms of interpolating rational quadratic (RQ) functions; these RQ samples are differentiable and have analytic inverses, so that they satisfy the properties required for coupled transformations. The NSF directly controls knots and derivatives by adjusting the residual network.

C. Network architecture

In the ongoing fourth LIGO-Virgo observing run (O4), a total of four detectors are in operation. Given the temporal correlation between signals from different detectors, the data dimension is substantial, leading to some redundancy. Consequently, a prevalent approach for data processing (feature extraction) involves the utilization of multi-input machine learning [90]. In this context, we employed the front-end residual net (ResNet) to extract features from the data d .

In this section, we outline the architecture of the ResNet-50 [124] and employed the normalizing flow model. Two distinct networks, namely 1D ResNet-50 and 2D ResNet-50, were utilized to handle the input time strain and its spectrogram, respectively. Their outputs were then combined into a unified 1D vector.

ResNet-50, a 50-layer deep network, incorporates a unique feature known as skip connections. These connections link the output of one layer to another by bypassing intermediate layers. Combined with these skip connections, the layers form residual blocks, enhancing the appropriateness of the initial weights by enabling the network to learn residuals, with the output typically approaching 0. The activation function for each residual block is rectified linear units (ReLU), effectively capturing complex representations and addressing the vanishing gradient problem.

An innovative aspect of our approach involves the addition of a dropout layer with a value of 0.2 after each activation function within the ResNet. This strategic inclusion mitigates the risk of overfitting and diminishes the

Table 2. The architecture of the ResNet model for time-domain strain data in the two-detector network. The first column shows the name of the used layer. The second column shows the dimensions of the output data obtained by the corresponding layer. The third column shows the specific structure of each layer, with labels a and b denoting the one-dimensional residual block. Here, each filter size is denoted by a , and the output channel is represented by b . “stride c ” means that the stride of the convolution or pooling layer is c .

Layer name	Output size	Architecture
Input layer	(2048, 2)	–
Conv1	(1024, 64)	7, 64, stride 2
MaxPooling	(512, 64)	3 max pool, stride 2
Conv2_x	(256, 256)	$\begin{bmatrix} 1, 64 \\ 3, 64 \\ 1, 256 \end{bmatrix} \times 3$
Conv3_x	(128, 512)	$\begin{bmatrix} 1, 128 \\ 3, 128 \\ 1, 512 \end{bmatrix} \times 3$
Conv4_x	(64, 1024)	$\begin{bmatrix} 1, 256 \\ 3, 256 \\ 1, 1024 \end{bmatrix} \times 3$
Conv5_x	(32, 2048)	$\begin{bmatrix} 1, 512 \\ 3, 512 \\ 1, 2048 \end{bmatrix} \times 3$
Flatten	(2048)	–

Table 3. Same as Table 2, but the filter size is $a \times a$.

Layer name	Output size	Architecture
Input layer	(200, 200, 2)	–
Conv1	(100, 100, 64)	Cov2D(7×7, 64), stride 2
MaxPooling	(50, 50, 64)	3×3 max pool, stride 2
Conv2_x	(25, 25, 256)	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
Conv3_x	(13, 13, 512)	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 3$
Conv4_x	(7, 7, 1024)	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 3$
Conv5_x	(4, 4, 2048)	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
Flatten	(2048)	–

network's reliance on input features, potentially enhancing its suitability for processing data containing glitches.

For the subsequent normalizing flow model, we em-

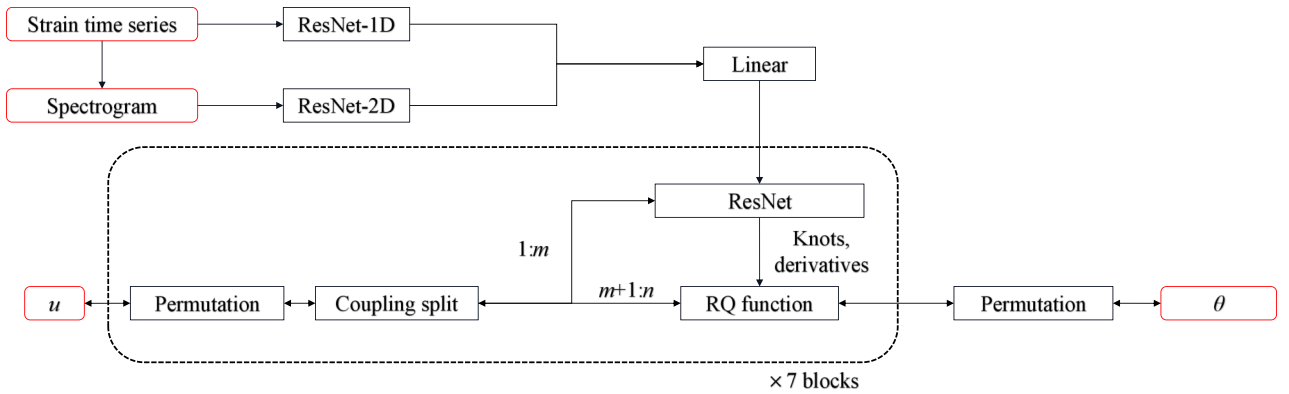


Fig. 4. (color online) The workflow of TTSF-NF. The input data comprises 2-second time strain data and a spectrogram. The data traverse through the different ResNet-50 network constructed using 1D convolutional layers and 2D convolutional layers, respectively. The features extracted from the two ResNet-50 models are subsequently merged into a 1D feature vector. This feature vector is then employed as a conditional input for the normalizing flow, generating samples from the base distribution and transforming them into the posterior distribution.

ployed the ReLU activation function with the hidden layer sizes set to 4096, 9 flow steps, 7 transform blocks, and 8 bins.

In this study, our focus centered on testing events resembling GW150914, which is a representative event. Consequently, we constructed a two-detector network, although the network's applicability extends to scenarios with four or more detectors. The specific structure of the front-end ResNet is detailed in Tables 2 and 3, while the network's specific architecture is illustrated in Fig. 4. In cases in which only the time strain or spectrogram was input for comparison, the corresponding network was omitted accordingly. The network employed in this study was implemented using PyTorch [125] and Lampe [126]. Figures were generated using Matplotlib [127] and ligo.skymap [128].

Throughout the training process, we iteratively generated data by sampling the prior distributions of the events and obtaining a new noise realization for each data frame. This approach serves to augment the number of training sets and minimizes the risk of NPE producing overly confident or excessively conservative posterior distributions [129]. The AdamW optimizer [130] was employed with a learning rate set to 0.0001, a batch size of 200, and a learning rate decay factor of 0.99. For initializing the network parameters with initial random values, the "Xavier" initialization [131] was applied. This initialization method aims to maintain an appropriate scale for the weights during both forward and backward propagation, mitigating issues such as gradient disappearance or explosion. Its design principle involves initializing weights to random values that adhere to a specific distribution, ensuring consistency in the input and output variances.

Training incorporated an early stopping strategy to achieve convergence and prevent overfitting. The network underwent approximately 16 days of training on a single NVIDIA GeForce RTX A6000 GPU with 48 GB

of memory.

III. RESULTS AND DISCUSSION

A. Results and reliability

Before drawing specific inferences, it is crucial to assess the reliability of the method's results. This is achieved by conducting the Kolmogorov-Smirnov (KS) test [132] to compare the one-dimensional posterior distributions generated by the TTSF-NF outputs. Taking blips as an example, Hanford's O3b data comprises a total of 1437 entries with SNR > 12, and 1300 of these were selected for training. Subsequently, KS testing was per-

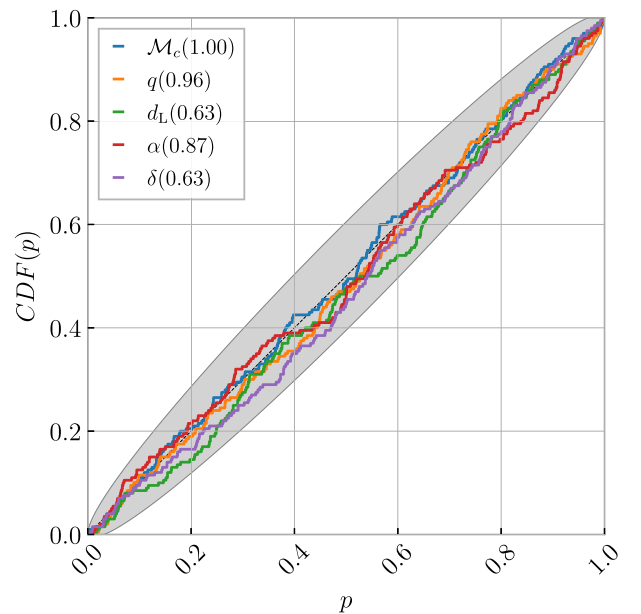


Fig. 5. (color online) P-P plot for 200 simulated datasets containing blips analyzed by TTSF-NF.

formed using 100 noise-injected simulated waveforms containing blips, which were not part of the training process. Figure 5 illustrates the construction of a probability-probability (P-P) plot based on 200 simulated datasets. For each parameter, we calculated the percentile value of the true value within its marginal posterior, plotting the cumulative distribution function (CDF) of this value, which represents the range of intervals covered by the corresponding confidence interval. In a well-behaved posterior, the percentiles should be uniformly distributed, resulting in a diagonal CDF. The p -values for the KS test are provided in the figure legends, and the gray area denotes the 3σ confidence limit. The proximity of the CDF to the diagonal indicates the model's ability to accurately sample the posterior, confirming the reasonability of the parameter range given by this method.

Figure 6 displays the posterior distribution of an event, emphasizing our innovative approach to exploring non-Gaussian noise and proposing new solution ideas. To initially evaluate the effectiveness of this method, we compared it with the posterior distributions in the absence of glitches. Position parameter inference results from Bilby [133] and the network are shown for scenarios with and without blips in the noise. The interval time chosen for this comparison, following Macas *et al.* [42],

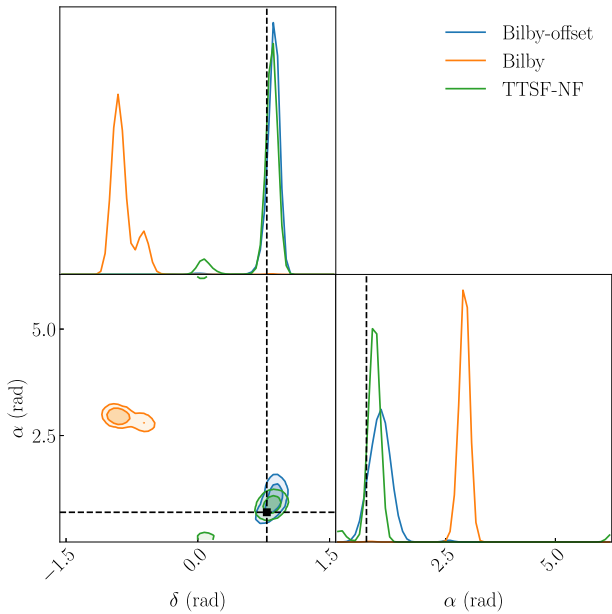


Fig. 6. (color online) One- and two-dimensional marginalized posterior distributions for δ and α . The intersection points of the dashed lines are the injected parameters. Here, the Bilby-offset represents the posterior parameters predicted by Bilby with noise (without blip) plus the GW signal. Bilby represents the posterior parameters predicted by Bilby with noise (with blip) plus the GW signal. TTSF-NF represents the posterior parameters predicted by the network with noise (with blip) plus the GW signal.

is $(t_0 - 30)$ ms, when the blip has the most significant impact on positioning. The results indicate that TTSF-NF can avoid position errors induced by glitch-contaminated data and achieves an accuracy comparable to Bilby with the dynesty sampler [134] on glitch-free data. Although we observe a bimodal shape in the network's posterior distribution, the probability of the second peak is low and can be disregarded. This phenomenon may arise from the narrow width of ResNet and the relatively simple feature extraction [135].

B. Comparative analysis

We proceeded to investigate whether it is necessary to input both the time strain and the spectrogram. Initially, we examined the overall loss of the network, as depicted in Fig. 7. The loss comparison suggests that TTSF-NF exhibits a better performance, although T-NF still performs reasonably well. However, these losses are based on the 1300 blips in the training set. To ascertain if the network has genuinely learned the characteristics of blips, we evaluated its generalization performance on the remaining 100 glitches that were not part of the training set.

For glitches not included in the training set, TTSF-NF outperforms T-NF. The overall parameter estimation performance is quantified by the determination coefficient R^2 under the same source parameters, defined as follows:

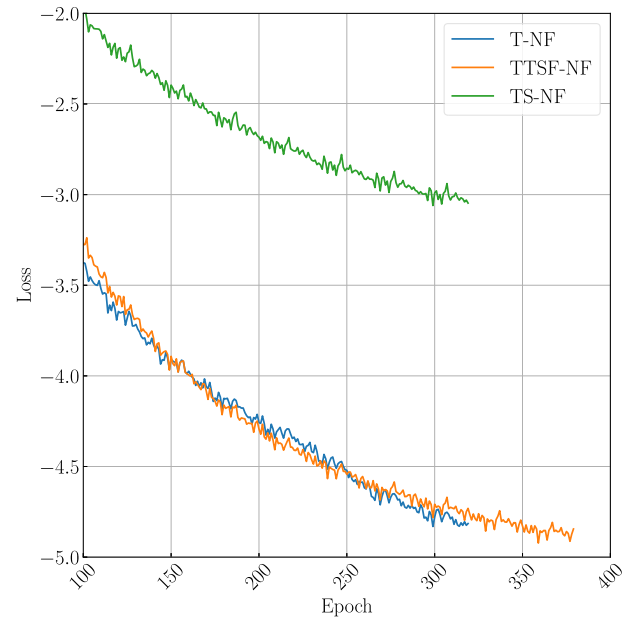


Fig. 7. (color online) The loss of all different networks for each epoch. During the training process, we used 10 cycles of early stops to ensure that each network could reach the optimal. Here, T-NF represents the loss of the network using only time strain. TTSF-NF represents the loss of the network using both time strain and a spectrogram. TF-NF represents the loss of the network using only a spectrogram.

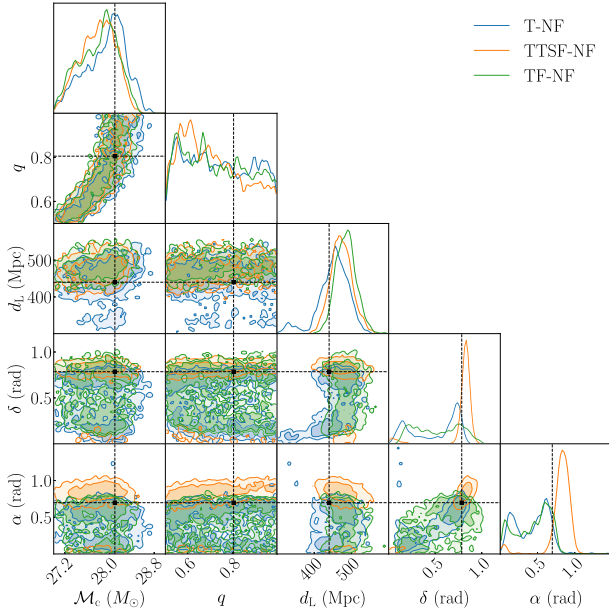


Fig. 8. (color online) One- and two-dimensional marginalization of M_c , q , d_L , δ and α with blip for different NF estimates of the posterior distribution. The intersection points of the dashed lines are the injected parameters. Here, T-NF represents the posterior parameters predicted by the network using only time strain. TTSF-NF represents the posterior parameters predicted by the network using both time strain and a spectrogram. TF-NF represents the posterior parameters predicted by the network using only a spectrogram.

$$R^2 = 1 - \frac{\sum (y_{\text{pre}} - y_{\text{true}})^2}{\sum (y_{\text{pre}} - \bar{y}_{\text{true}})^2}, \quad (6)$$

where y_{pre} and y_{true} represent the variable estimates in the sample and the true value of the test sample, respectively, and \bar{y}_{true} represents the average true value of the test sample. We calculated R^2 for 200 test datasets in our preset range for both networks. For T-NF, $R^2 = 0.66$; for TTSF-NF, $R^2 = 0.71$. Therefore, incorporating both time strain and a spectrogram yields significant improvements.

Figure 8 presents the specific posterior distributions, notably in α and δ . The time-spectrogram normalizing

flow (TS-NF) reveals the widest posterior distribution, which is attributed to limitations in the time-frequency resolution and binning that may result in the loss of detailed information. Additionally, the posterior distribution is not significantly enhanced for T-NF. This is because the glitch's impact on the GW parameters primarily stems from temporal and frequency overlap. Consequently, TSF-NF struggles to differentiate between signal and noise information. TTSF-NF combines the strengths of both approaches. It emphasizes the contributions of different frequency components in the time series, facilitating a clearer distinction between signal and noise, while preserving time resolution as much as possible. This makes it more suitable for signal processing when the channel is contaminated by glitches. Despite the potential longer computation time required for spectrogram calculation using the Q -transform, the benefits of leveraging both representations outweigh the potential delay, particularly when considering the enhanced analytical capabilities they provide.

The results indicate that the posterior distributions of each NF for the luminosity distance d_L , chirp mass M_c , and mass ratio q are similar, which might be attributed to glitch extraction. However, it is premature to conclude that glitches affect different parameters inconsistently across neural networks. Table 4 presents the specific posterior distributions after data cleansing. Our findings reveal that the network performed best when utilizing both time strain data and spectrograms, achieving an accuracy comparable to that for glitch-free data. Notably, while the network inference's posterior distribution accuracy exceeds that of the Bilby inference, this discrepancy could stem from the non-stationary nature of the noise at varying time instances.

The processing time for each network and the comparison with Bilby are outlined in Table 5. For the spectrogram-applied method, we considered the data processing time utilizing 32 CPUs in parallel. It is evident that regardless of the method used, the processing time is significantly faster than that of the traditional method. When combined with the Q -transform time, the overall

Table 4. A comparison between previously injected parameters and parameters recovered by Bilby and normalizing flow methods. The recovered values are accompanied by their 2σ confidence regions. Here, the Bilby-offset represents the posterior parameters predicted by Bilby with noise (without blip) plus GW signal. Bilby represents the posterior parameters predicted by Bilby using noise (with blip) plus the GW signal.

Parameter	Injected value	Bilby-offset	Bilby	T-NF	TTSF-NF	TS-NF
$M_c (M_\odot)$	28.10	$27.85^{+0.28}_{-0.47}$	$28.28^{+0.35}_{-0.53}$	$27.97^{+0.55}_{-0.86}$	$27.79^{+0.52}_{-0.66}$	$27.55^{+0.53}_{-0.76}$
q	0.83	$0.72^{+0.26}_{-0.18}$	$0.67^{+0.29}_{-0.15}$	$0.72^{+0.26}_{-0.21}$	$0.69^{+0.29}_{-0.18}$	$0.71^{+0.28}_{-0.20}$
$d_L (\text{Mpc})$	440.00	$423.57^{+58.27}_{-38.89}$	$426.26^{+27.59}_{-25.04}$	$453.89^{+64.94}_{-125.52}$	$468.75^{+60.76}_{-52.80}$	$456.7^{+63.11}_{-63.11}$
α (rad)	0.69	$0.84^{+0.15}_{-0.33}$	$2.89^{+0.08}_{-0.16}$	$0.46^{+0.28}_{-0.42}$	$0.83^{+0.15}_{-0.16}$	$0.36^{+0.37}_{-0.58}$
δ (rad)	0.78	$0.83^{+0.03}_{-0.13}$	$-0.92^{+0.32}_{-0.10}$	$0.63^{+0.17}_{-0.53}$	$0.82^{+0.07}_{-0.08}$	$0.52^{+0.40}_{-0.39}$

Table 5. A comparison of the computational time required by different methods for generating their respective samples. Note that the input data is noise (with blip) plus GW signal.

Sampling Method	Number of posterior samples	Total runtime/s	Time per sample/s
Bilby	2111	246.41	0.1165
T-NF	1024	0.66	0.0006
TTSF-NF	1024	1.35	0.0013
TS-NF	1024	1.13	0.0011

processing time is not significantly longer, indicating the continued suitability of the Q -transform. In practical applications, additional time for calibration, strain data distribution, and signal identification must be considered. The time required for these tasks averages around 0.4 s for 2 s of time strain data [136, 137]. All networks can complete processing within 2 s (the length of the data), making them suitable for real-time data processing.

C. Other glitches

Given that various glitch types exhibit different distributions in time and frequency, whereby blips are typically short-term, it is crucial to demonstrate the algorithm's suitability for other glitch forms. To investigate this aspect, we chose scattered light noise for validation, aiming to determine if the method is effective for long-duration glitches. We utilized 1200 glitches for both training and testing.

Figure 9 presents the P-P plot generated from 200 simulated datasets for testing. This plot underscores the reliability of the posterior distribution produced by our network, even when the noise includes scattered noise. Specific posterior distribution details are illustrated in Fig. 10.

IV. CONCLUSION

GW detectors often encounter disruptive non-Gaussian noise artifacts known as glitches. The occurrence of these glitches in proximity to GW events can significantly impact subsequent parameter estimates. Conventional deglitching methods, while effective, demand substantial computational resources, posing challenges for achieving real-time deglitching at higher frequencies of GW events in the future. This, in turn, could impede timely observations of physical phenomena, including EM counterpart observations.

In this study, we leverage TTSF-NF to expedite parameter inference when GW data are marred by glitches. This assumes critical significance for swift wave source localization and real-time analysis, particularly as the frequency of future events escalates. Our pioneering approach involves combining the high temporal resolution

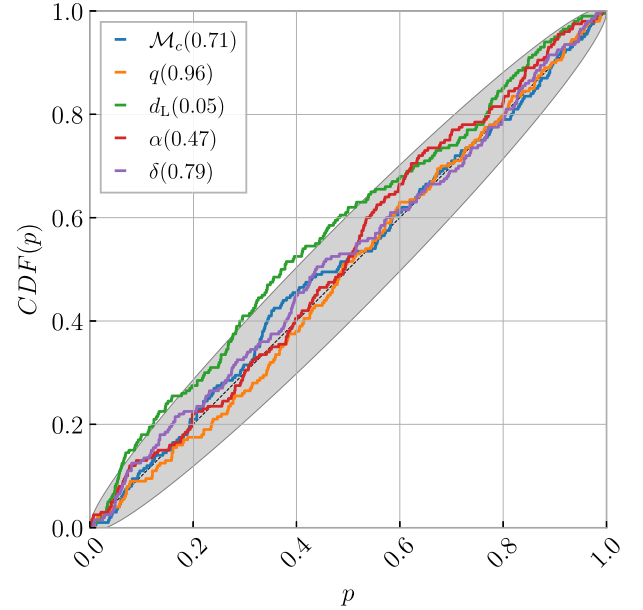


Fig. 9. (color online) P-P plot for 200 simulated datasets containing scattered light noises analyzed by TTSF-NF.

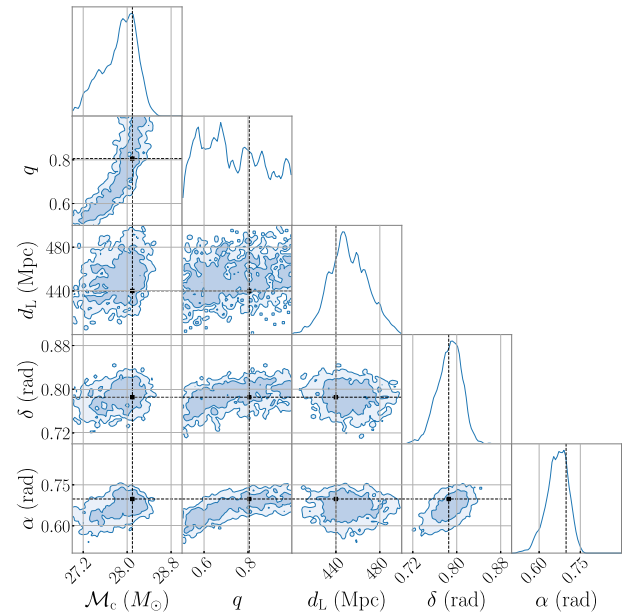


Fig. 10. (color online) One- and two-dimensional marginalized posterior distributions for M_c , q , d_L , δ and α with scattered light noise. The intersection points of the dashed lines are the injected parameters.

of the time domain with the distinct discriminability of features in the time-frequency domain, aiming for rapid and judicious parameter inference in glitch-contaminated data. Notably, our choice of the normalizing flow as a more flexible flow contributes to the success of this innovative methodology.

Specifically, our focus was on glitches with a SNR > 12 , which represents one of the most prevalent

glitch types in GW detectors that existing robust methods struggle to effectively process. Notably, we discovered that relying solely on the spectrogram for parameter inference is suboptimal due to resolution limitations. Although the use of only the time strain on the training set produced effects equivalent to utilizing both time strain and a spectrogram, the network relying solely on time strain struggled to effectively discriminate between features in the time domain for previously unseen glitches. The incomplete separation of these features resulted in an inferior performance compared with the network utilizing both time strain and a spectrogram. Our proposed method achieves real-time data processing, processing 2 s of data in 1.35 s. Additionally, we verified the applicability of this model to scattered light noise.

The integration of the normalizing flow opens promising avenues for the future real-time processing of glitch-contaminated data. It is essential to note that the

upper limit of the frequency in this study is 1024 Hz, and as the frequency of BNS mergers exceeds this threshold, with BNS remaining in the detector's sensitive range for longer than 2 s, direct migration of this network to BNS scenarios is not viable. In our subsequent work, we plan to address the BNS scenario and explore more suitable front-end network structures for better data fusion. Simultaneously, we will consider implementing methods such as Cohen's Class of Time-Frequency Representations to enhance the time-frequency domain resolution for an optimal performance.

ACKNOWLEDGEMENTS

This research has made use of data or software obtained from the Gravitational Wave Open Science Center (gwosc.org), a service of LIGO Laboratory, the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. We thank He Wang for helpful discussions.

References

- [1] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. Lett.* **116**, 061102 (2016), arXiv:1602.03837[gr-qc]
- [2] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. Lett.* **119**, 161101 (2017), arXiv:1710.05832[gr-qc]
- [3] B. P. Abbott *et al.* (LIGO Scientific, Virgo, Fermi-GBM, INTEGRAL), *Astrophys. J. Lett.* **848**, L13 (2017), arXiv:1710.05834[astro-ph.HE]
- [4] B. P. Abbott *et al.* (LIGO Scientific, Virgo, Fermi GBM, INTEGRAL, IceCube, AstroSat Cadmium Zinc Telluride Imager Team, IPN, Insight-Hxmt, ANTARES, Swift, AGILE Team, 1M2H Team, Dark Energy Camera GW-EM, DES, DLT40, GRAWITA, Fermi-LAT, ATCA, ASKAP, Las Cumbres Observa-tory Group, OzGrav, DWF (Deeper Wider Faster Program), AST3, CAASTRO, VINROUGE, MAS-TER, J-GEM, GROWTH, JAGWAR, CaltechNRAO, TTU-NRAO, NuSTAR, Pan-STARRS, MAXI Team, TZAC Consortium, KU, Nordic Optical Telescope, ePESSTO, GROND, Texas Tech University, SALT Group, TOROS, BOOTES, MWA, CALET, IKI-GW Follow-up, H.E.S.S., LOFAR, LWA, HAWC, Pierre Auger, ALMA, Euro VLBI Team, Pi of Sky, Chan-dra Team at McGill University, DFN, ATLAS Tele-scopes, High Time Resolution Universe Survey, RIMAS, RATIR, SKA South Africa/MeerKAT), *Astrophys. J. Lett.* **848**, L12 (2017), arXiv:1710.05833[astro-ph.HE]
- [5] J. Aasi *et al.* (LIGO Scientific), *Class. Quant. Grav.* **32**, 074001 (2015), arXiv:1411.4547[gr-qc]
- [6] F. Acernese *et al.* (VIRGO), *Class. Quant. Grav.* **32**, 024001 (2015), arXiv:1408.3978[gr-qc]
- [7] T. Akutsu *et al.* (KAGRA), *Nature Astron.* **3**, 35 (2019), arXiv:1811.08079[gr-qc]
- [8] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. Lett.* **116**, 221101 (2016) [Erratum: *Phys. Rev. Lett.* **121**, 129902 (2018)], arXiv: 1602.03841[gr-qc]
- [9] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. Lett.* **123**, 011102 (2019), arXiv:1811.00364[gr-qc]
- [10] R. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. D* **103**, 122002 (2021), arXiv:2010.14529[gr-qc]
- [11] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), (2021), arXiv: 2112.06861[gr-qc]
- [12] M. Isi, M. Giesler, W. M. Farr *et al.*, *Phys. Rev. Lett.* **123**, 111102 (2019), arXiv:1905.00869[gr-qc]
- [13] N. Yunes, K. Yagi, and F. Pretorius, *Phys. Rev. D* **94**, 084002 (2016), arXiv:1603.08955[gr-qc]
- [14] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Phys. Rev. Lett.* **121**, 161101 (2018), arXiv:1805.11581[gr-qc]
- [15] E. Annala, T. Gorda, A. Kurkela *et al.*, *Phys. Rev. Lett.* **120**, 172703 (2018), arXiv:1711.02644[astro-ph.HE]
- [16] B. Margalit and B. D. Metzger, *Astrophys. J. Lett.* **850**, L19 (2017), arXiv:1710.05938[astro-ph.HE]
- [17] B. P. Abbott *et al.* (LIGO Scientific, Virgo), *Astrophys. J. Lett.* **818**, L22 (2016), arXiv:1602.03846[astro-ph.HE]
- [18] I. Mandel and F. S. Broekgaarden, *Living Rev. Rel.* **25**, 1 (2022), arXiv:2107.14239[astro-ph.HE]
- [19] F. S. Broekgaarden *et al.*, *Monthly Notices of the Royal Astronomical Society* **516**(4), 5737 (2022), arXiv:2112.05763[astro-ph.HE]
- [20] B. P. Abbott *et al.* (LIGO Scientific, Virgo, 1M2H, Dark Energy Camera GW-E, DES, DLT40, Las Cumbres Observatory, VINROUGE, MASTER), *Nature* **551**, 85 (2017), arXiv:1710.05835[astro-ph.CO]
- [21] S. Bird, I. Cholis, J. B. Muñoz *et al.*, *Phys. Rev. Lett.* **116**, 201301 (2016), arXiv:1603.00464[astro-ph.CO]
- [22] J. Sakstein and B. Jain, *Phys. Rev. Lett.* **119**, 251303 (2017), arXiv:1710.05893[astro-ph.CO]
- [23] L.-F. Wang, X.-N. Zhang, J.-F. Zhang *et al.*, *Phys. Lett. B* **782**, 87 (2018), arXiv:1802.04720[astro-ph.CO]
- [24] X.-N. Zhang, L.-F. Wang, J.-F. Zhang *et al.*, *Phys. Rev. D* **99**, 063510 (2019), arXiv:1804.08379[astro-ph.CO]
- [25] X. Zhang, *Sci. China Phys. Mech. Astron.* **62**, 110431 (2019), arXiv:1905.11122[astro-ph.CO]
- [26] J.-F. Zhang, H.-Y. Dong, J.-Z. Qi *et al.*, *Eur. Phys. J. C* **80**, 217 (2020), arXiv:1906.07504[astro-ph.CO]
- [27] L.-F. Wang, Z.-W. Zhao, J.-F. Zhang *et al.*, *JCAP* **11**, 012

- (2020), arXiv:1907.01838[astro-ph.CO]
- [28] J.-F. Zhang, M. Zhang, S.-J. Jin *et al.*, *JCAP* **09**, 068 (2019), arXiv:1907.03238[astro-ph.CO]
- [29] Z.-W. Zhao, L.-F. Wang, J.-F. Zhang *et al.*, *Sci. Bull.* **65**, 1340 (2020), arXiv:1912.11629[astro-ph.CO]
- [30] S.-J. Jin, D.-Z. He, Y. Xu *et al.*, *JCAP* **03**, 051 (2020), arXiv:2001.05393[astro-ph.CO]
- [31] L.-F. Wang, S.-J. Jin, J.-F. Zhang *et al.*, *Sci. China Phys. Mech. Astron.* **65**, 210411 (2022), arXiv:2101.11882[gr-qc]
- [32] S.-J. Jin, L.-F. Wang, P.-J. Wu *et al.*, *Phys. Rev. D* **104**, 103507 (2021), arXiv:2106.01859[astro-ph.CO]
- [33] J.-Y. Song, L.-F. Wang, Y. Li *et al.*, *Sci. China Phys. Mech. Astron.* **67**, 230411 (2022)
- [34] T.-N. Li, S.-J. Jin, H.-L. Li *et al.*, *Sci. China Phys. Mech. Astron.* **963**, 1 (2024)
- [35] S.-J. Jin, Y.-Z. Zhang, J.-Y. Song *et al.*, (2023), arXiv:2305.19714[astro-ph.CO]
- [36] L. Barack *et al.*, *Class. Quant. Grav.* **36**, 143001 (2019), arXiv:1806.05195[gr-qc]
- [37] A. Palmese, C. R. Bom, S. Mucesh *et al.*, *Astrophys. J.* **943**, 56 (2023), arXiv:2111.06445[astro-ph.CO]
- [38] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), *Phys. Rev. X* **9**, 031040 (2019), arXiv:1811.12907[astro-ph.HE]
- [39] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), *Phys. Rev. X* **11**, 021053 (2021), arXiv:2010.14527[gr-qc]
- [40] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), *Phys. Rev. X* **13**, 041039 (2023)
- [41] B. P. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration), *Class. Quant. Grav.* **35**, 065010 (2018), arXiv:1710.02185[gr-qc]
- [42] R. Macas, J. Pooley, L. K. Nuttall *et al.*, *Phys. Rev. D* **105**, 103021 (2022), arXiv:2202.00344[astro-ph.HE]
- [43] B. P. Abbott *et al.*, *Classical and Quantum Gravity* **37**, 055002 (2020)
- [44] T. D. Canton, S. Bhagwat, S. V. Dhurandhar *et al.*, *Classical and Quantum Gravity* **31**, 015016 (2014)
- [45] D. Davis *et al.* (LIGO), *Class. Quant. Grav.* **38**, 135014 (2021), arXiv:2101.11673[astro-ph.IM]
- [46] J. Powell, *Class. Quant. Grav.* **35**, 155017 (2018), arXiv:1803.11346[astro-ph.IM]
- [47] S. Biscans, J. Warner, R. Mittleman *et al.*, *Classical and Quantum Gravity* **35**, 055004 (2018)
- [48] J. Glanzer, S. Soni, J. Spoon *et al.*, *Class. Quant. Grav.* **40**, 195015 (2023), arXiv:2304.07477[astro-ph.IM]
- [49] A. Buikema *et al.* (aLIGO), *Phys. Rev. D* **102**, 062003 (2020), arXiv:2008.01301[astro-ph.IM]
- [50] P. Nguyen *et al.* (AdvLIGO), *Class. Quant. Grav.* **38**, 145001 (2021), arXiv:2101.09935[astro-ph.IM]
- [51] E. Schwartz *et al.*, *Classical and Quantum Gravity* **37**, 235007 (2020)
- [52] A. F. Brooks *et al.* (LIGO Scientific), *Appl. Opt.* **60**, 4047 (2021), arXiv:2101.05828[physics.ins-det]
- [53] M. Punturo *et al.*, *Class. Quant. Grav.* **27**, 194002 (2010)
- [54] M. Dominik, E. Berti, R. O'Shaughnessy *et al.*, *Astrophys. J.* **806**, 263 (2015), arXiv:1405.7016[astro-ph.HE]
- [55] B. P. Abbott *et al.* (LIGO Scientific), *Class. Quant. Grav.* **34**, 044001 (2017), arXiv:1607.08697[astro-ph.IM]
- [56] S. Soni *et al.*, *Class. Quant. Grav.* **38**, 195016 (2021), arXiv:2103.12104[gr-qc]
- [57] C. Pankow *et al.*, *Phys. Rev. D* **98**, 084016 (2018), arXiv:1808.03619[gr-qc]
- [58] M. Hannam *et al.*, *Nature* **610**, 652 (2022), arXiv:2112.11300[gr-qc]
- [59] E. Payne, S. Hourihane, J. Golomb *et al.*, *Phys. Rev. D* **106**, 104017 (2022), arXiv:2206.11932[gr-qc]
- [60] N. J. Cornish and T. B. Littenberg, *Class. Quant. Grav.* **32**, 135012 (2015), arXiv:1410.3835[gr-qc]
- [61] N. J. Cornish, T. B. Littenberg, B. Bécsy *et al.*, *Phys. Rev. D* **103**, 044006 (2021), arXiv:2011.09494[gr-qc]
- [62] D. Davis, T. J. Massinger, A. P. Lundgren *et al.*, *Class. Quant. Grav.* **36**, 055011 (2019), arXiv:1809.05348[astro-ph.IM]
- [63] D. Davis, T. B. Littenberg, I. M. Romero-Shaw *et al.*, *Class. Quant. Grav.* **39**, 245013 (2022), arXiv:2207.03429[astro-ph.IM]
- [64] N. J. Cornish, *Phys. Rev. D* **103**, 104057 (2021), arXiv:2101.01188[gr-qc]
- [65] B. Steltner, M. A. Papa, and H.-B. Eggenstein, *Phys. Rev. D* **105**, 022005 (2022), arXiv:2105.09933[gr-qc]
- [66] C. Talbot, E. Thrane, S. Biscoveanu *et al.*, *Phys. Rev. Res.* **3**, 043049 (2021), arXiv:2106.13785[astro-ph.IM]
- [67] S. Soni *et al.*, *Class. Quant. Grav.* **38**, 025016 (2020), arXiv:2007.14876[astro-ph.IM]
- [68] S. Hourihane, K. Chatziioannou, M. Wijngaarden *et al.*, *Phys. Rev. D* **106**, 042006 (2022), arXiv:2205.13580[gr-qc]
- [69] S. D. Mohanty and M. A. T. Chowdhury, *Class. Quant. Grav.* **40**, 125001 (2023), arXiv:2301.02398[gr-qc]
- [70] R. Udall and D. Davis, *Appl. Phys. Lett.* **122**, 094103 (2023), arXiv:2211.15867[astro-ph.IM]
- [71] S. A. Usman *et al.*, *Class. Quant. Grav.* **33**, 215004 (2016), arXiv:1508.02357[gr-qc]
- [72] C. Messick *et al.*, *Phys. Rev. D* **95**, 042001 (2017), arXiv:1604.04324[astro-ph.IM]
- [73] M. I. Jordan and T. M. Mitchell, *Science* **349**, 255 (2015)
- [74] W. Wei and E. A. Huerta, *Phys. Lett. B* **800**, 135081 (2020), arXiv:1901.00869[gr-qc]
- [75] Z. Ren, H. Wang, Y. Zhou *et al.*, (2022), arXiv:2212.14283[gr-qc]
- [76] S.-J. Jin, Y.-X. Wang, T.-Y. Sun *et al.*, (2023), arXiv:2305.19003[gr-qc]
- [77] C. Chatterjee, L. Wen, F. Diakogiannis *et al.*, *Phys. Rev. D* **104**, 064046 (2021), arXiv:2105.03073[gr-qc]
- [78] R. Biswas *et al.*, *Phys. Rev. D* **88**, 062003 (2013), arXiv:1303.6984[astro-ph.IM]
- [79] H. Yu and R. X. Adhikari, *Front. Artif. Intell.* **5**, 811563 (2022), arXiv:2111.03295[astro-ph.IM]
- [80] G. Vajente, Y. Huang, M. Isi *et al.*, *Phys. Rev. D* **101**, 042003 (2020), arXiv:1911.09083[gr-qc]
- [81] R. Ormiston, T. Nguyen, M. Coughlin *et al.*, *Phys. Rev. Res.* **2**, 033066 (2020), arXiv:2005.06534[astro-ph.IM]
- [82] J. Merritt, B. Farr, R. Hur *et al.*, *Phys. Rev. D* **104**, 102004 (2021), arXiv:2108.12044[gr-qc]
- [83] G. Ashton, *Monthly Notices of the Royal Astronomical Society* **520**, 2983 (2023)
- [84] S. Bini, G. Vedovato, M. Drago *et al.*, *Class. Quant. Grav.* **40**, 135008 (2023), arXiv:2303.05986[gr-qc]
- [85] K. Mogushi, R. Quitzow-James, M. Cavaglia *et al.*, *Mach. Learn. Sci. Tech.* **2**, 035018 (2021), arXiv:2101.04712[gr-qc]
- [86] D. Rezende and S. Mohamed, in *International conference on machine learning* (PMLR, 2015) pp. 1530–1538.

- [87] G. Papamakarios and I. Murray, in *Advances in Neural Information Processing Systems*, Vol. 29, edited by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Curran Associates, Inc., 2016).
- [88] J.-M. Lueckmann, P. J. Goncalves, G. Bassetto *et al.*, *Advances in Neural Information Processing Systems* **30**, (2017)
- [89] D. Greenberg, M. Nonnenmacher, and J. Macke, in *International Conference on Machine Learning* (PMLR, 2019) pp. 2404–2414.
- [90] M. Dax, S. R. Green, J. Gair *et al.*, *Phys. Rev. Lett.* **127**, 241103 (2021), arXiv:2106.12594[gr-qc]
- [91] H. Gabbard, C. Messenger, I. S. Heng *et al.*, *Nature Phys.* **18**, 112 (2022), arXiv:1909.06296[astro-ph.IM]
- [92] A. J. K. Chua and M. Vallisneri, *Phys. Rev. Lett.* **124**, 041102 (2020), arXiv:1909.05966[gr-qc]
- [93] S. Hassan, F. Villaescusa-Navarro, B. Wandelt *et al.*, *The Astrophysical Journal* **937**, 83 (2022)
- [94] X. Zhao, Y. Mao, C. Cheng *et al.*, *Astrophys. J.* **926**, 151 (2022), arXiv:2105.03344[astro-ph.CO]
- [95] X. Zhao, Y. Mao, and B. D. Wandelt, *Astrophys. J.* **933**, 236 (2022), arXiv:2203.15734[astro-ph.CO]
- [96] X. Zhao, Y. Mao, S. Zuo *et al.*, (2023), arXiv:2310.17602[astro-ph.IM]
- [97] D. George, H. Shen, and E. A. Huerta, *Phys. Rev. D* **97**, 101501 (2018)
- [98] N. Lopac, F. Hri, I. P. Vuksanovi *et al.*, *IEEE Access* **10**, 2408 (2022)
- [99] D. George and E. Huerta, *Phys. Lett. B* **778**, 64 (2018)
- [100] L. Yang and S. Hong, in *International Conference on Machine Learning* (PMLR, 2022) pp. 25038–25054.
- [101] Z. Liu, Q. Ma, P. Ma *et al.*, in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37 (2023) pp. 8923–8931.
- [102] W. Zhang, J. Zhang, J. Li *et al.*, in *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management* (2023) pp. 3308–3318.
- [103] S. Coughlin, M. Zevin, S. Bahaadini *et al.*, (2021)
- [104] M. Du, B. Liang, H. Wang *et al.*, (2023), arXiv:2308.05510[astro-ph.IM]
- [105] M. Cabero *et al.*, *Class. Quant. Grav.* **36**, 15 (2019), arXiv:1901.05093[physics.ins-det]
- [106] A. H. Nitz, *Class. Quant. Grav.* **35**, 035016 (2018), arXiv:1709.08974[gr-qc]
- [107] L. Nuttall *et al.*, *Class. Quant. Grav.* **32**, 245005 (2015), arXiv:1508.07316[gr-qc]
- [108] T. Accadia *et al.*, *Class. Quant. Grav.* **27**, 194011 (2010)
- [109] T. D. Gebhard, N. Kilbertus, I. Harry *et al.*, *Phys. Rev. D* **100**, 063015 (2019), arXiv:1904.08693[astro-ph.IM]
- [110] P. Nousi, A. E. Koloniari, N. Passalis *et al.*, *Phys. Rev. D* **108**, 024022 (2023), arXiv:2211.01520[gr-qc]
- [111] S. Alvarez-Lopez, A. Liyanage, J. Ding *et al.*, (2023), arXiv:2304.09977[gr-qc]
- [112] A. Trovato *et al.*, *POS Proceedings of Science* **82**, (2019)
- [113] S. Chatterji, L. Blackburn, G. Martin *et al.*, *Class. Quant. Grav.* **21**, S1809 (2004), arXiv:gr-qc/0412119
- [114] C. M. Biwer, C. D. Capano, S. De *et al.*, *Publ. Astron. Soc. Pac.* **131**, 024503 (2019), arXiv:1807.10312[astro-ph.IM]
- [115] S. Khan, K. Chatziioannou, M. Hannam *et al.*, *Phys. Rev. D* **100**, 024059 (2019), arXiv:1809.10113[gr-qc]
- [116] S. R. Green and J. Gair, *Mach. Learn. Sci. Tech.* **2**, 03L (2021), arXiv:2008.03312[astro-ph.IM]
- [117] M. J. Williams, J. Veitch, and C. Messenger, *Phys. Rev. D* **103**, 103006 (2021), arXiv:2102.11056[gr-qc]
- [118] J. Langendorff, A. Kolmus, J. Janquart *et al.*, *Phys. Rev. Lett.* **130**, 171402 (2023), arXiv:2211.15097[gr-qc]
- [119] D. Ruhe, K. Wong, M. Cranmer *et al.*, (2022), arXiv:2211.09008[astro-ph.IM]
- [120] W. Ruan, H. Wang, C. Liu *et al.*, *Universe* **9**, 407 (2023), arXiv:2307.14844[astro-ph.IM]
- [121] I. Csizsar, *The Annals of Probability* **3**, 146 (1975)
- [122] G. Papamakarios, D. Sterratt, and I. Murray, in *The 22nd International Conference on Artificial Intelligence and Statistics* (PMLR, 2019) pp. 837–848.
- [123] C. Durkan, A. Bekasov, I. Murray *et al.*, *Advances in Neural Information Processing Systems* **32**, (2019)
- [124] K. He, X. Zhang, S. Ren *et al.*, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016) pp. 770–778.
- [125] A. Paszke, S. Gross, F. Massa, A. Lerer *et al.*, *Advances in Neural Information Processing Systems* **32**, (2019) <https://github.com/francois-rozet/lampe>
- [126] J. D. Hunter, *Computing in Science & Engineering* **9**, 90 (2007)
- [127] <https://lscsoft.docs.ligo.org/ligo.skymap/> ().
- [128] J. Hermans, A. Delaunoy, F. Rozet *et al.*, *stat.* **1050**, (2022)
- [129] I. Loshchilov and F. Hutter, (2018)
- [130] S. K. Kumar, (2017), arXiv: 1704.08863
- [131] J. Veitch *et al.*, *Phys. Rev. D* **91**, 042003 (2015), arXiv:1409.7215[gr-qc]
- [132] G. Ashton *et al.*, *Astrophys. J. Suppl.* **241**, 27 (2019), arXiv:1811.02042[astro-ph.IM]
- [133] J. S. Speagle, *Mon. Not. Roy. Astron. Soc.* **493**, 3132 (2020), arXiv:1904.02180[astro-ph.IM]
- [134] Q. Ailing and T. Ning, in *2020 2nd International Conference on Information Technology and Computer Application (ITCA)* (2020) pp. 588–592.
- [135] D. George and E. A. Huerta, *Phys. Rev. D* **97**, 044039 (2018), arXiv:1701.00008[astro-ph.IM]
- [136] A. Viets *et al.*, *Class. Quant. Grav.* **35**, 095015 (2018), arXiv:1710.09973[astro-ph.IM]