ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/180459/

This is the author’s version of a work that was submitted to / accepted for publication.

Citation for final published version:

Raymond, Vivien , Al-Shammari, Sama and Gottel, Alexandre 2025. Simulation-based inference for gravitational-waves from intermediate-mass binary black holes in real noise. Monthly Notices of the Royal Astronomical Society 10.1093/mnras/staf1282

Publishers page: https://dx.doi.org/10.1093/mnras/staf1282

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher’s version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

Simulation-based Inference for Gravitational-waves from Intermediate-Mass Binary Black Holes in Real Noise

Vivien Raymond,[1]★ Sama Al-Shammari,[1] Alexandre Göttel[1]
[1]*School of Physics and Astronomy, Cardiff University, Cardiff, CF24 3AA, United Kingdom*

Accepted 2025 July 30. Received 2025 July 30; in original form 2024 June 25

ABSTRACT

We present an exploratory investigation into using Simulation-based Inference techniques, specifically Flow-Matching Posterior Estimation, to construct a posterior density estimator trained using real gravitational-wave detector noise. Our prototype estimator is trained on a 9-dimensional space, and for training efficiency outputs posterior probability distributions for the binary black holes chirp mass and mass ratio. We use this prototype estimator to investigate possible effects on parameter estimation for Intermediate-Mass Binary Black Holes, and show statistically significant reduction in measurement bias. Although the results show potential for improved measurements, they also highlight the need for further work.

Key words: gravitational waves – methods: data analysis

1 INTRODUCTION

The LIGO Buikema et al. (2020), Virgo Acernese et al. (2019) and KAGRA Akutsu et al. (2020) Gravitational Wave (GW) observatories have since 2014 reported around one hundred GWs from binary black hole mergers, with individual black hole masses ranging from a few stellar masses to over a hundred Abbott et al. (2023). The heaviest detected so far lie at the edge of the regime associated with Intermediate-Mass Black Hole (IMBH), above about one hundred solar masses. As outlined in Greene et al. (2020), these black holes could play a fundamental role in stellar and galactic evolution, and represent a significant source of gravitational waves and tidal disruption events. However, in the detection band of current ground-based gravitational-wave detectors, IMBH signals are typically only detectable in very short time intervals of a few milliseconds. Thus, short duration noise artefacts (known as glitches) present in the noise of those detectors have the potential to strongly affect the measurements of IMBHs. During LIGO's third observing run, glitches with a Signal-to-Noise Ratio (SNR) of at least 6.5 occurred about once per minute Abbott et al. (2023), with the actual rate of relevant glitches being higher due to the potential impact of smaller glitches on parameter estimation.

The parameter estimation methods commonly used by the LIGO-Virgo-KAGRA (LVK) collaboration, such as those available in the Bilby library Ashton et al. (2019), have proven very reliable for inferring black-hole parameters. However, they generally assume that the noise around events is purely Gaussian and stationary. When glitches or other deviations occur, these assumptions can significantly impact the results Powell (2018); Davis et al. (2022); Hourihane et al. (2022); Macas et al. (2022); Ashton (2023). Simulation-Based Inference (SBI) on the other hand, leverages machine learning to infer parameters without assuming a specific noise distribution, provided sufficient training data are available Cranmer et al. (2020). Previous studies have applied SBI and Neural Posterior Estimation (NPE) towards GW parameter estimation Green et al. (2020); Delaunoy et al. (2020); Dax et al. (2021, 2022, 2023); Bhardwaj et al. (2023); Alvey et al. (2024) using Gaussian stationary noise. In particular, Dax et al. (2023) used Flow Matching Posterior Estimation (FMPE) on gravitational-wave signals in Gaussian noise. The work presented here applied the same technique, on a more limited gravitational-wave signal space, but looking at the impact of real, non-Gaussian and non-stationary detector noise. Notably, Wildberger et al. (2023) has been able to address Power Spectral Density (PSD) uncertainties, while Legin et al. (2023b); Legin et al. (2023a) employ Score-Based Likelihood Characterization to create a likelihood function based on real detector noise.

In this work, we use SBI to directly map simulated IMBH signals in real detector noise to the posterior distributions of inferred black hole parameters, for training efficiency we currently only infer the binary's chirp mass and mass ratio. This is the first effort to train fully amortized networks for parameter estimation on real detector noise, eliminating the need for both importance sampling and likelihood assumptions. This approach allows us to realistically study the effects of real detector noise while leveraging the speed of SBI.

This paper is organised as follows: Section 2 describes our SBI methods and simulations, Section 3.1 presents results on signals generated in Gaussian noise and Section 3.2 does the same on real detector noise. Finally, Section 4 summarizes our findings and discusses them in the context of future developments.

2 METHODS

The field of GW parameter inference currently relies on Bayesian sampling methods to retrieve astrophysical information from GW

★ E-mail: raymondv@cardiff.ac.uk

signals. These methods are all based on Bayes' theorem:

$$p(\theta|x) = \frac{p(x|\theta) \cdot p(\theta)}{p(x)}, \tag{1}$$

where $p(\theta|x)$ is the posterior probability of the parameters θ given the observed data x, $p(x|\theta)$ is the likelihood of the data x given the parameters θ, $p(\theta)$ is the prior of the parameters θ before observation, and $p(x)$ is the marginal likelihood or evidence *i.e.* the probability of observing the data under all considered parameter values. It is calculated as:

$$p(x) = \int p(x|\theta)p(\theta)\,d\theta \tag{2}$$

SBI is a class of Bayesian machine learning methods that utilise simulated data in order to infer probability density distributions. In this paper we use Likelihood-free AMortized Posterior Estimation (LAMPE)'s Rozet et al. (2021) implementation of FMPE Lipman et al. (2022); Dax et al. (2023) as our density estimator. To train the neural network, we require only mechanistic models (in our case these are the GW waveform models), constraints on the prior and segments of real on-site detector noise. We sample from the prior and in conjunction with our models and noise segments, we simulate synthetic data $x \sim p(x|\theta)$ to give as input to a normalising flows neural network Rezende & Mohamed (2015); Papamakarios et al. (2021). Normalising flows define a probability distribution $q(\theta|x)$ over n number of parameters in the parameter space $\theta \in \mathbf{R}^n$ in terms of an invertible mapping $\psi_x : \mathbf{R}^n \rightarrow \mathbf{R}^n$ from a simple base distribution $q_0(\theta)$ Rezende & Mohamed (2015); Papamakarios et al. (2021):

$$q(\theta|x) = (\psi_x)_* q_0(\theta) = q_0(\psi_x^{-1}(\theta)) \left| \det \frac{\partial \psi_x^{-1}(\theta)}{\partial \theta} \right|, \tag{3}$$

where $(\psi)_*$ denotes the forward flow operator, and x is conditioned as $x \in \mathbf{R}^m$, where m is the dimension of the observed data space, i.e., how many data points or features are in each observation x. Normalizing flows are discrete flows, such that ψ_x is a collection of simpler mappings with triangular Jacobians and θ shuffling. This results in a neural density estimator, $q(\theta|x)$, that is simple to evaluate, quick to sample from and approximates the posterior $p(\theta|x)$. Flow matching is a method that uses a vector field v_t to directly define the velocity of sample trajectories as they move towards the target distribution Lipman et al. (2022). These trajectories are determined by solving ordinary differential equations (ODEs), which allows flow matching to achieve optimal transport without the need for discrete diffusion paths. This means that flow matching can directly reach the target distribution and compute densities more efficiently than other generative methods, such as Neural Posterior Score Estimation (NPSE) Sharrock et al. (2022); Geffner et al. (2023); Sohl-Dickstein et al. (2015); Song & Ermon (2019); Ho et al. (2020). FMPE is a technique that applies flow matching to Bayesian inference Lipman et al. (2022); Dax et al. (2023), it works by directly aligning the estimated posterior distribution with the true posterior distribution. This alignment is achieved through a loss function that minimizes the difference between the two distributions. Due to this and the continuous nature of the flow, FMPE can lead to a more direct and possibly more accurate estimation of the posterior as opposed to other methods such as Sequential Neural Posterior Estimation (SNPE)Rezende & Mohamed (2015); Papamakarios et al. (2021); Bhardwaj et al. (2023). In those methods the posterior distribution is iteratively refined through sequential updates with a heavy reliance on approximations and intermediate layers, or NPSE where they focus on estimating the score function (gradient of the log-posterior) rather than the posterior distribution itself, leading to intractable posterior densities.

In the FMPE regime, we utilise continuous normalising flows, which are parameterised by a continuous "time" parameter $t \in [0, 1]$, such that $q_{t=0}(\theta|x) = q_0(\theta)$ and $q_{t=1}(\theta|x) = q(\theta|x)$ Chen et al. (2018). Each t defines the flow by a vector field $v_{t,x}$ on the sample space. This corresponds to the velocity of the sample trajectories,

$$\frac{d}{dt}\psi_{t,x}(\theta) = v_{t,x}(\psi_{t,x}(\theta)), \quad \psi_{0,x}(\theta) = \theta. \tag{4}$$

Integrating this ODE then gives the trajectories $\theta_t \equiv \psi_{t,x}(\theta)$. The final density is retrieved by solving the transport equation $\partial_t q_t + div(q_t v_{t,x}) = 0$ and is:

$$q(\theta|x) = (\psi_{1,x})_* q_0(\theta) = q_0(\theta) \exp\left(-\int_0^1 \operatorname{div} v_{t,x}(\theta_t)\, dt \right). \tag{5}$$

The continuous flow thus allows $v_{t,x}(\theta)$ to be specified simply by a neural network taking $\mathbf{R}^{n+m+1} \rightarrow \mathbf{R}^n$. The main goal of flow matching training is to make the learned vector field $v_{t,x}$ closely follow a target vector field $u_{t,x}$. This target vector field generates a path $p_{t,x}$ that leads us towards the posterior distribution we want to estimate. By doing this, we avoid the need to solve ordinary differential equations (ODEs) during training. Although choosing the pair $(u_{t,x}, p_{t,x})$ might seem complex initially, Lipman et al. (2022) showed that the training process becomes much simpler if we condition the path on θ_1, a sample drawn from the prior distribution, instead of x. This is known as sample-conditional basis. For a given sample-conditional probability path $p_t(\theta|\theta_1)$ with a corresponding vector field $u_t(\theta|\theta_1)$, the sample-conditional flow matching loss is defined as

$$L_{\mathrm{SCFM}} = \mathbb{E}_{t\sim U[0,1], x\sim p(x), \theta_1\sim p(\theta|x), \theta_t\sim p_t(\theta_t|\theta_1)} \cdot \left[\|v_{t,x}(\theta_t) - u_t(\theta_t|\theta_1)\|^2 \right]. \tag{6}$$

According to Lipman et al. (2022), minimising this loss is equivalent to regressing $v_{t,x}(\theta)$ on the marginal vector field $u_{t,x}$ that generates $p_t(\theta|x)$. Due to the sample-conditional vector field being independent of x the $x-$dependence of $v_{t,x}(\theta)$ is picked up by the expectation value. Flow matching is applied to SBI by using Bayes' theorem to make the replacement $\mathbf{E}_{\mathbf{p}(\mathbf{x})\mathbf{p}(\theta|\mathbf{x})} \rightarrow \mathbf{E}_{\mathbf{p}(\theta)\mathbf{p}(\mathbf{x}|\theta)}$, removing the intractable expectation values, making the new FMPE loss:

$$L_{\mathrm{FMPE}} = \mathbb{E}_{t\sim p(t), \theta_1\sim p(\theta), x\sim p(x|\theta_1), \theta_t\sim p_t(\theta_t|\theta_1)} \cdot \left[\|v_{t,x}(\theta_t) - u_t(\theta_t|\theta_1)\|^2 \right]. \tag{7}$$

We generalise the uniform distribution in Equation (6) by sampling from $t \sim p(t), t \in [0, 1]$ in this expression as well. This provides more freedom to improve learning in our networks.

The family of Gaussian sample-conditional paths are first presented in Lipman et al. (2022) and are given as

$$p_t(\theta|\theta_1) = \mathcal{N}(\theta|\mu_t(\theta_1), \sigma_t(\theta_1)^2 \mathbf{I}_n) \tag{8}$$

where one could freely specify, contingent on boundary conditions, the time-dependant means $\mu_t(\theta_1)$ and standard deviations $\sigma_t(\theta_1)$. The sample-conditional probability path must be chosen to concentrate around θ_1 at $t = 1$ (within a small region of size σ_{min}) in addition to being the base distribution at $t = 0$. In this work, we utilise the optimal transport paths (shown in Lipman et al. (2022) and used in Dax et al. (2023)) defined by $\mu_t(\theta_1)$ and $\sigma_t(\theta_1) = 1 - (1 - \sigma_{\min})t$ making the sample-conditional vector field have the form

$$u_t(\theta|\theta_1) = \frac{\theta_1 - (1 - \sigma_{\min})\theta}{1 - (1 - \sigma_{\min})t}. \tag{9}$$

Training data is generated by sampling from θ from the prior

[ht]

Parameter	**Range**	**Prior**
Chirp mass	80-120 $M_\odot$	Uniform
Mass ratio	0.3-1.0	Uniform
Luminosity distance	1-4 Gpc	Uniform
Time	20 ms wide	Uniform

Table 1. Prior parameters. All other parameters are using the standard uniform priors Abbott et al. (2023).

and in conjunction with waveform models and detector source noise, simulating data x corresponding to θ. The FMPE loss in Equation (7) is minimised via empirical risk minimisation over samples $(\theta, x) \sim p(\theta)p(x|\theta)$.

Generative diffusion or flow matching models typically handle complex, high-dimensional data (such as images) in the θ space. They often use U-Net architectures to map θ to a vector field $v(\theta)$ of the same dimension, with t and an optional conditioning vector x included. In the SBI case however, and particularly in this study field, the data is often complicated whereas the parameters θ are low dimensional. This indicates that it would be more useful in our case to build the network architecture as a mapping that goes from x to $v(x)$ and then apply conditioning on θ and t. We can therefore use any feature extraction architecture for the data and in our case we use Singular Value Decompositions (SVD)s to extract the most informative features of the data segments. Note that SVDs built on the GW models may not always useful because they can remove relevant features from the noise.

3 RESULTS

The results presented here were created using networks trained with simulated data from non-spinning IMBH models using the IMRPhenomXPHM waveform approximant Pratten et al. (2021), generated from the 9-dimensional parameter space, injected in different kinds of noise in a single GW detector (the LIGO Hanford detector). Furthermore, during training only the chirp-mass $\mathcal{M}$ and mass-ratio q were labelled, thus limiting network's output to those two parameters. This effectively marginalises over all the other parameters (sky location, distance, phase, polarization, merger time, and inclination angle), making training easier. Time and sky-location parameters local to the (single) detector were used instead of the standard geocenter time, right-ascension and declination parameters. This enables faster training while still allowing us to investigate the effect of real noise on the inference.

For this exploratory investigation the priors used are listed in Table 1. The high masses allow for a segment length of 0.5 s and a high frequency cutoff of 256 Hz, while the low frequency cutoff is set to the usual 20 Hz, relevant for the Advanced LIGO zero-detuned high power Abbott et al. (2020a) noise curve used to generate Gaussian noise. We do highlight that for GW190521 which has similar mass parameters as our injection for this exploratory investigation, previous Bayesian analyses have gone below this threshold Abbott et al. (2020b), while others have also used 20 Hz, e.g. Abbott et al. (2024).

This work uses LAMPE's implementation of FMPE which we combine with GWpy Macleod et al. (2021) and Bilby library Ashton et al. (2019); Romero-Shaw et al. (2020a). In particular, we are using a Multi-Layer Perceptron with 32 layers of 256 features and Exponential Linear Unit activation function. Despite the risks of vanishing gradients, those hyper parameters did provide the best results over a sweep from 2^{3-6} layers and 2^{6-10} features. Training was done using about 10 million segments in the frequency domain, each including a simulated waveform randomly drawn from the prior, taking about 1 day on a A100 GPU.

3.1 Gaussian noise

When trained on injections with Gaussian noise, we expect for the SBI network to converge towards a representation of the conditional posterior mapping able to directly sample from the posterior with the correspondingly estimated Gaussian likelihood. To check the convergence of the training we perform a set of simulated injections in Gaussian noise in the LIGO Hanford detector using IMRPhenomXPHM and recover them with both the trained SBI network, as well as the Dynesty Speagle (2020) sampler implemented in the bilby software library. We used the same Advanced LIGO zero-detuned high power Abbott et al. (2020a) noise curve and the IMRPhenomXPHM waveform model. We ran 4 parallel analyses using 4000 live points each. Two examples are shown in the top panels of Figure 1, where the sampled distributions are near-identical with Jensen-Shannon Divergence (JSD) values of 0.001 nat and 0.006 nat, respectively, thus confirming that our training samples accurately represented the 9-dimensional parameter space. We provide the JSD values as they are commonly used in gravitational-wave astronomy Abbott et al. (2019). A typical threshold to indicate statistical agreement is 0.035 nat.

Furthermore, a 2-dimensional percentile-percentile test (Figure 2, panel (a)) on 100 injections recovered with the trained network shows that our credible intervals behave as expected, which matches the performance from stochastic sampling, see for instance Romero-Shaw et al. (2020b). We calculated the two-dimensional mass coverage with a Monte Carlo integration Hermans et al. (2022) for this plot.

3.2 LIGO detector noise

The training data was generated by injecting simulated signals at randomly selected starting points within a roughly 14 h stretch of LIGO Hanford data around February 20^{th} 2020 (GPS time 1266213786.7). For training, we used the same network architecture from Section 3.1. In most cases this reproduced the results from the Gaussian likelihood as sampled with Dynesty using bilby which estimated the PSD using the default median average method settings Romero-Shaw et al. (2020a), other settings being identical to those in the previous section. For example, the results in the bottom right panel of Figure 1 reached a JSD of 0.004 nat and 0.013 nat for the mass ratio q and chirp-mass $\mathcal{M}$, respectively.

However, for a fraction of injections, by not assuming Gaussianity nor stationarity, the network outperformed the Gaussian likelihood, as shown in Figure 1, bottom left panel, with a JSD of 0.008 nat for the mass ratio q, and 0.07 nat for the chirp-mass $\mathcal{M}$.

The statistical reliability of the credible interval of this network was assessed with a percentile-percentile plot (see Figure 2 panel (b)), which gave the expected diagonal behaviour. Note that the network is by construction specific to the data used in training, and as such is only amortized over observations in noise with similar properties. In practice it may be necessary to retrain the network if the detector characterizations analyses find differences.

4 DISCUSSION

The improvement achieved by modelling the real noise distribution in our results is comparable to the Score-Based Likelihood Characterization findings of Legin et al. (2023a). However, unlike their

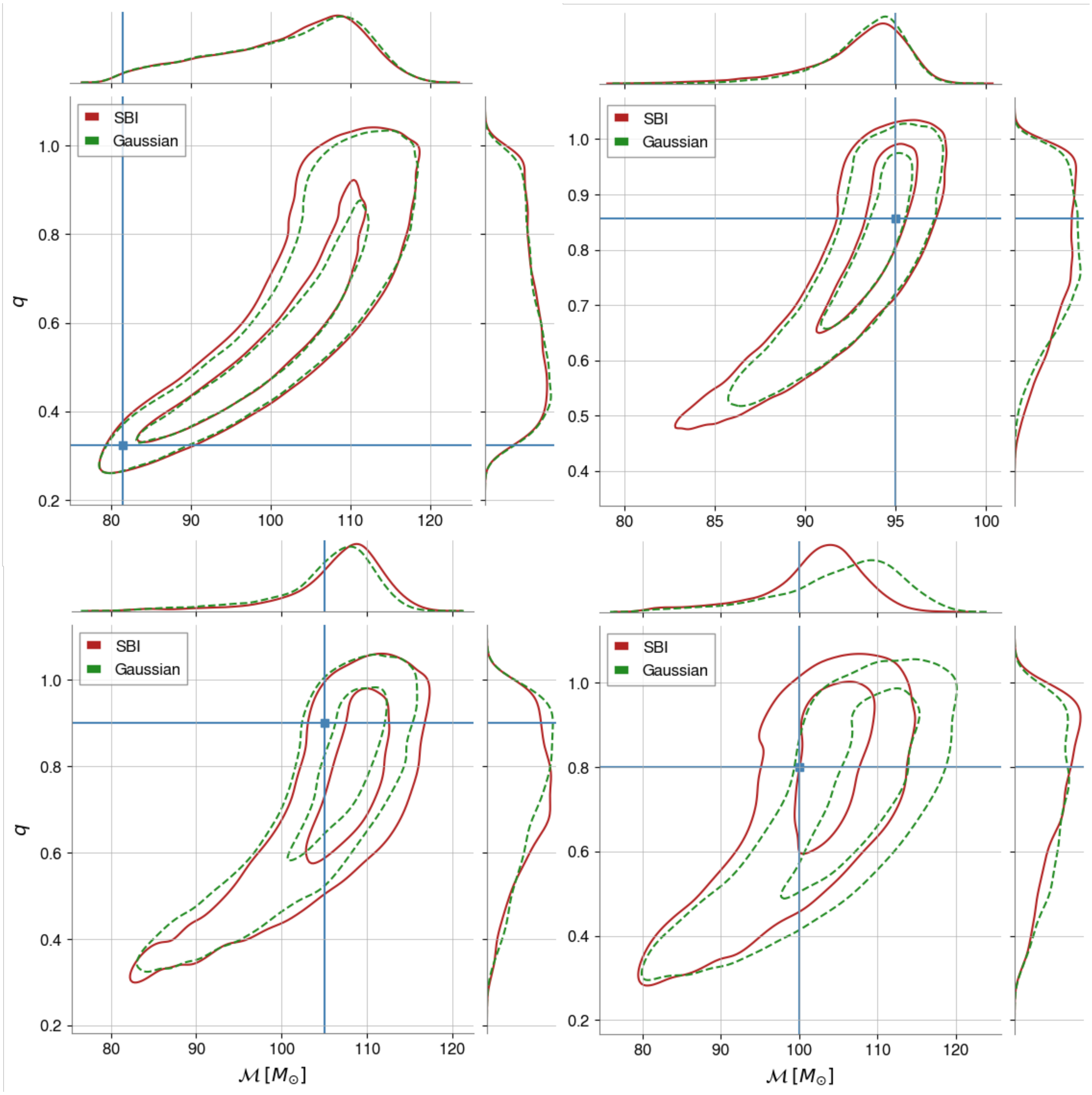

Figure 1. PDFs for the chirp-mass $\mathcal{M}$ and the mass-ratio q from sampling a Gaussian Likelihood and a SBI network trained with (top panels) Gaussian noise on an injection into Gaussian noise, and trained with (bottom panels) real detector noise on an injection into real detector noise. The blue lines show the injected values.

approach, our trained SBI network directly generates samples from the target conditional posterior distribution and does not require an extra sampling step. On this limited example, after a ≈ 24 h training on a A100 GPU, the bottleneck for inference time shifts to our GPU's I/O capacities, reducing it to a few milliseconds, whereas standard sampling requires several hours.

Our results thus represent a first look at how simulation-based inference may be able to perform optimal parameter estimation using real detector noise. Current limitations include the need to characterize and quantify noise features beyond Gaussianity and stationarity. We emphasise however that this works lays a strong foundation and can in future investigations be used to focus on precessing-spin analyses that also include calibration error modeling. Testing on simulated analytical non-Gaussian noise distributions, and testing with different PSD estimation methods such as BayesWave Cornish et al. (2021), will be able to lead us to robust SBI architectures for GW inference.

Beyond the extension to precessing systems, future studies will also involve increasing the domain of applicability, specifically low-

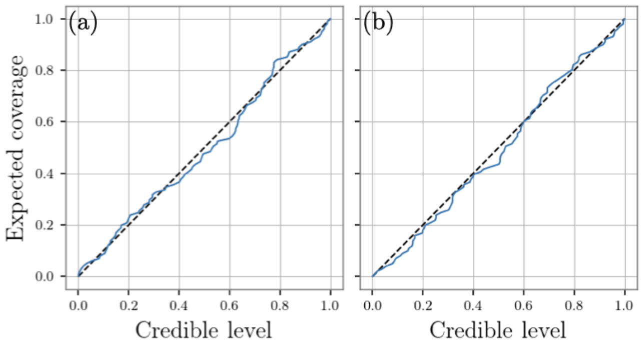

Figure 2. Percentile-percentile plots on the multi-dimensional PDF sampled from the converged neural network trained on a) Gaussian noise (p-value 0.64) and b) real detector noise (p-value 0.66), using 100 injections.

ering the mass range on which the network is trained, as currently the analysis would not be reliable on lower-mass signals, and including multiple detectors and eccentric signals. However, given our SVD compression pre-processing step, and the performance of other compression techniques on GW signals, for instance Reduced-Order-Modelling methods (see Tiglio & Villanueva (2022) for a review), the scaling is expected to be manageable. Additionally, work on marginalising over multiple waveform approximants such that the resulting marginalised posterior distribution encompasses errors from waveform systematics and will help towards enlarging the domain of applicability of the networks.

ACKNOWLEDGEMENTS

The authors would like to thank Stephen Green, Maximilian Dax, Virginia d'Emilio and Alex Nitz for helpful discussions. This work was supported by the UK's Science and Technology Facilities Council grant ST/V005618/1, the Royal Society Award ICA\R1\231114 and the Leverhulme Trust Fellowship IF-2024-038. 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. The authors are grateful for computational resources provided by the LIGO Laboratory and Cardiff University and supported by National Science Foundation Grants PHY-0757058 and PHY-0823459, and STFC grants ST/I006285/1 and ST/V005618/1.

DATA AVAILABILITY

The data utilized for this work will be freely available upon reasonable request to the corresponding author.

REFERENCES

Abbott B. P., et al., 2019, Phys. Rev. X, 9, 031040
Abbott B. P., et al., 2020a, Living Reviews in Relativity, 23, 3
Abbott R., et al., 2020b, Astrophys. J. Lett., 900, L13
Abbott R., et al., 2023, Phys. Rev. X, 13, 041039
Abbott R., et al., 2024, Phys. Rev. D, 109, 022001
Acernese F., et al., 2019, Phys. Rev. Lett., 123, 231108
Akutsu T., et al., 2020, Progress of Theoretical and Experimental Physics, 2021, 05A101
Alvey J., Bhardwaj U., Domcke V., Pieroni M., Weniger C., 2024, Physical Review D, 109, 083008
Ashton G., 2023, Monthly Notices of the Royal Astronomical Society, 520, 2983
Ashton G., et al., 2019, The Astrophysical Journal Supplement Series, 241, 27
Bhardwaj U., Alvey J., Miller B. K., Nissanke S., Weniger C., 2023, ArXiV
Buikema A., et al., 2020, Phys. Rev. D, 102, 062003
Chen R. T., Rubanova Y., Bettencourt J., Duvenaud D. K., 2018, Advances in neural information processing systems, 31
Cornish N. J., Littenberg T. B., Bécsy B., Chatziioannou K., Clark J. A., Ghonge S., Millhouse M., 2021, Phys. Rev. D, 103, 044006
Cranmer K., Brehmer J., Louppe G., 2020, Proceedings of the National Academy of Sciences, 117, 30055
Davis D., Littenberg T. B., Romero-Shaw I. M., Millhouse M., McIver J., Di Renzo F., Ashton G., 2022, Class. Quantum Grav., 39, 245013
Dax M., Green S. R., Gair J., Macke J. H., Buonanno A., Schölkopf B., 2021, Physical review letters, 127, 241103
Dax M., Green S., Gair J., Deistler M., Schölkopf B., Macke J. H., 2022, in ICLR 2022.
Dax M., Wildberger J., Buchholz S., Green S. R., Macke J. H., Schölkopf B., 2023, arXiv preprint arXiv:2305.17161
Delaunoy A., Wehenkel A., Hinderer T., Nissanke S., Weniger C., Williamson A. R., Louppe G., 2020, arXiv preprint arXiv:2010.12931
Geffner T., Papamakarios G., Mnih A., 2023, in International Conference on Machine Learning. pp 11098–11116
Green S. R., Simpson C., Gair J., 2020, Physical Review D, 102, 104057
Greene J. E., Strader J., Ho L. C., 2020, Annual Review of Astronomy and Astrophysics, 58, 257
Hermans J., Delaunoy A., Rozet F., Wehenkel A., Begy V., Louppe G., 2022, A Trust Crisis In Simulation-Based Inference? Your Posterior Approximations Can Be Unfaithful (`arXiv:2110.06581`), `https://arxiv.org/abs/2110.06581`
Ho J., Jain A., Abbeel P., 2020, Advances in neural information processing systems, 33, 6840
Hourihane S., Chatziioannou K., Wijngaarden M., Davis D., Littenberg T., Cornish N., 2022, Phys. Rev. D, 106, 042006
Legin R., Wong K., Isi M., Adam A., Perreault-Levasseur L., Hezaveh Y., 2023a, in Machine Learning for Astrophysics. p. 17
Legin R., Hezaveh Y., Perreault-Levasseur L., Wandelt B., 2023b, The Astrophysical Journal, 943, 4
Lipman Y., Chen R. T., Ben-Hamu H., Nickel M., Le M., 2022, arXiv preprint arXiv:2210.02747
Macas R., et al., 2022, Phys. Rev. D, 105, 103021
Macleod D. M., Areeda J. S., Coughlin S. B., Massinger T. J., Urban A. L., 2021, SoftwareX, 13, 100657

Papamakarios G., Nalisnick E., Rezende D. J., Mohamed S., Lakshminarayanan B., 2021, Journal of Machine Learning Research, 22, 1

Powell J., 2018, Class. Quantum Grav., 35, 155017

Pratten G., et al., 2021, Physical Review D, 103

Rezende D., Mohamed S., 2015, in International conference on machine learning. pp 1530–1538

Romero-Shaw I. M., et al., 2020a, Mon. Not. Roy. Astron. Soc., 499, 3295

Romero-Shaw I. M., et al., 2020b, Mon. Not. Roy. Astron. Soc., 499, 3295

Rozet F., Delaunoy A., Miller B., et al., 2021, LAMPE: Likelihood-free Amortized Posterior Estimation, doi:10.5281/zenodo.8405782, `https://pypi.org/project/lampe`

Sharrock L., Simons J., Liu S., Beaumont M., 2022, arXiv preprint arXiv:2210.04872

Sohl-Dickstein J., Weiss E., Maheswaranathan N., Ganguli S., 2015, in International conference on machine learning. pp 2256–2265

Song Y., Ermon S., 2019, Advances in neural information processing systems, 32

Speagle J. S., 2020, Monthly Notices of the Royal Astronomical Society, 493, 3132

Tiglio M., Villanueva A., 2022, Living Reviews in Relativity, 25, 2

Wildberger J., Dax M., Green S. R., Gair J., Pürrer M., Macke J. H., Buonanno A., Schölkopf B., 2023, Physical Review D, 107, 084046