Lightning-Fast Gravitational Wave Parameter Inference through Neural Amortization

Arnaud Delaunoy, Antoine Wehenkel
Montefiore Institute,
University of Liège, Belgium

Tanja Hinderer
Institute for Theoretical Physics, Utrecht University, The Netherlands
Gravitation Astroparticle Physics Amsterdam (GRAPPA),
Institute for Theoretical Physics Amsterdam and Delta Institute for Theoretical Physics,
University of Amsterdam, The Netherlands

Samaya Nissanke, Christoph Weniger
Gravitation Astroparticle Physics Amsterdam (GRAPPA),
Institute for Theoretical Physics Amsterdam and Delta Institute for Theoretical Physics,
University of Amsterdam, The Netherlands

Andrew R. Williamson
Institute for Cosmology and Gravitation,
University of Portsmouth, UK

Gilles Louppe
Montefiore Institute,
University of Liège, Belgium

Abstract

Gravitational waves from compact binaries measured by the LIGO and Virgo detectors are routinely analyzed using Markov Chain Monte Carlo sampling algorithms. Because the evaluation of the likelihood function requires evaluating millions of waveform models that link between signal shapes and the source parameters, running Markov chains until convergence is typically expensive and requires days of computation. In this extended abstract, we provide a proof of concept that demonstrates how the latest advances in neural simulation-based inference can speed up the inference time by up to three orders of magnitude – from days to minutes – without impairing the performance. Our approach is based on a convolutional neural network modeling the likelihood-to-evidence ratio and entirely amortizes the computation of the posterior. We find that our model correctly estimates credible intervals for the parameters of simulated gravitational waves.

1 Introduction

Inferring the parameters of sources of gravitational-waves (GW) such as compact binaries detected by LIGO [1] and Virgo [6] is based on the evaluation of the Bayesian posterior probability of fifteen or more parameters that govern the shape of the signals [5]. Because the mapping between parameters and signals requires millions of waveform model evaluations, the analysis is computationally very expensive. Each likelihood evaluation requires generating the GW signal corresponding to a set of

source parameters and computing its noise-weighted correlation with detector data [5]. At present, accurate component-mass estimates only become available hours or even days after the detection of a binary black-hole coalescence. A full Markov-Chain Monte Carlo (MCMC) parameter estimation takes days to weeks to complete *for each single event*. The improvements in sensitivity of the GW detector network for the next observing run, scheduled to start in late 2021, will lead to unprecedentedly high detection rates of binary merger events of ~ 1 per week to $\sim 1/$ per day [4]. There is thus an urgent need to develop fast and efficient methods for parameter estimation. Furthermore, low-latency detection and characterization of GWs is essential to trigger multi-messenger time-sensitive searches in order to find electromagnetic and/or astroparticle counterparts to the GW signal [23].

Latest advances in deep learning offer an exciting direction of research towards fast detection and parameter estimation of GWs. For *signal detection*, several previous works [17, 22, 16, 14, 15] have framed the problem as an instance of supervised classification and showcased how to make use of convolutional neural networks (CNNs) to treat GW signals. Results presented across these work have shown fast and accurate detection both for synthetic and real data, hence confirming that CNNs are a useful and promising tool to produce real-time triggers. For *parameter estimation*, deep learning-based alternatives to sampling algorithms have also been investigated. A large panels of neural architectures has been used, e.g. Bayesian Neural Networks [24], Conditional Variational-Autoencoders [13], Mixture Density Networks [9] and Normalizing Flows [18, 19] (for a recent review see Ref. [12]). These studies all point towards the fact that fast and accurate parameter estimation is within reach.

Here, we build upon previous recent developments [11, 21, 8] for simulation-based inference [10], and demonstrate that they can drastically accelerate Bayesian posterior parameter estimation for gravitational wave (GW) signals in realistic multi-detector scenarios. Our approach makes use of supervised classification models, which unlocks battle-tested neural network architectures for high-dimensional data such as CNNs, complementary to previous works. We find that, *e.g.*, the localization of binary black-hole mergers can be accelerated by at least three orders of magnitude without compromising precision or accuracy.

2 GW signal inference

The GW emission by the quasi-circular inspiral and merger of a binary black hole (BH) system may be characterized by eight parameters intrinsic to the system: the two BH masses $m_{1,2}$ and the six components of the spin vectors $\boldsymbol{S}_{1,2}$. Additionally, there are seven extrinsic parameters that relate the source position and orientation to the observer frame: the luminosity distance d_L, the right ascension α and the declination δ defining the position on the sky, the coalescence time t_c, the binary inclination angle θ_{JN}, the coalescence phase angle Φ and the polarization angle Ψ [5].

Here we use the waveform model `IMRPhenomPv2` [20], which includes a phenomenological frequency domain description of the inspiral, merger and final BH ringdown phases. For a proof of concept, we choose to use this model because it is very fast to evaluate while still incorporating precession effects. However the model complexity is not really a problem here because our method only requires the simulator once, to generate the training dataset. The simulation outputs the strain observed at each detector $\mathbf{h}(t, \boldsymbol{\vartheta})$, a vector function of the time and of the 15 waveform parameters, $\boldsymbol{\vartheta} \equiv \{m_1, m_2, \boldsymbol{S}_1, \boldsymbol{S}_2, d_L, \alpha, \delta, t_c, \theta_{\text{JN}}, \Phi, \Psi\}$. We denote synthetic and real time-series data by $\boldsymbol{x} = \mathbf{h} + \mathbf{n}$, where $\mathbf{n}$ refers to heteroscedastic Gaussian noise that depends on the detectors.

The task of gravitational wave parameter estimation is to determine the waveform model parameters $\boldsymbol{\vartheta}$ that are the most compatible with the strain data time series $\boldsymbol{x}$. In gravitational wave analysis, this inference problem is usually framed as the computation of the Bayesian posterior distribution

$$p(\boldsymbol{\vartheta}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta})}{p(\boldsymbol{x})}, \tag{1}$$

where $p(\boldsymbol{x}|\boldsymbol{\vartheta})$ is the likelihood of the model parameters $\boldsymbol{\vartheta}$ given the observed data time series $\boldsymbol{x}$, $p(\boldsymbol{\vartheta})$ is a prior distribution over the model parameters and $p(\boldsymbol{x})$ is the model evidence.

Since both the likelihood function and the prior can be evaluated, sampling algorithms such as MCMC or Nested sampling are commonly used to approximate the posterior [25]. Obtaining a sufficient number of posterior samples may however be computationally expensive, taking days for binary black-hole mergers, and weeks for binary neutron-star mergers.

3 Inference amortization

We build upon previous simulation-based inference algorithms [11, 21, 8] to approximate the likelihood-to-evidence ratio

$$r(\boldsymbol{x}|\boldsymbol{\vartheta}) \equiv \frac{p(\boldsymbol{x}|\boldsymbol{\vartheta})}{p(\boldsymbol{x})} \tag{2}$$

with a neural network. As demonstrated in [21], this can be achieved by considering as positive examples (labeled $y = 1$) those strain-parameter pairs $\boldsymbol{x}, \boldsymbol{\vartheta} \sim p(\boldsymbol{x}, \boldsymbol{\vartheta}) = p(\boldsymbol{x}|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta})$ which are distributed jointly, and as negative examples (labeled $y = 0$) those strain-parameter pairs $\boldsymbol{x}, \boldsymbol{\vartheta} \sim p(\boldsymbol{x})p(\boldsymbol{\vartheta})$ which are sampled marginally (independently from their respective marginal distributions). Under mild assumptions, the decision function modeled by the Bayes optimal classifier for this binary classification problem is

$$s^*(\boldsymbol{x}, \boldsymbol{\vartheta}) = \frac{p(\boldsymbol{x}, \boldsymbol{\vartheta})}{p(\boldsymbol{x}, \boldsymbol{\vartheta}) + p(\boldsymbol{x})p(\boldsymbol{\vartheta})}, \tag{3}$$

which recovers the likelihood-to-evidence ratio as

$$\frac{s^*(\boldsymbol{x}, \boldsymbol{\vartheta})}{1 - s^*(\boldsymbol{x}, \boldsymbol{\vartheta})} = \frac{p(\boldsymbol{x}, \boldsymbol{\vartheta})}{p(\boldsymbol{x})p(\boldsymbol{\vartheta})} = \frac{p(\boldsymbol{x}|\boldsymbol{\vartheta})}{p(\boldsymbol{x})} = r(\boldsymbol{x}|\boldsymbol{\vartheta}). \tag{4}$$

After an upfront simulation and training phase, Bayesian inference for any observed strain $\boldsymbol{x}$ is *amortized*: the fast evaluation of the posterior reduces to

$$\hat{p}(\boldsymbol{\vartheta}|\boldsymbol{x}) = \hat{r}(\boldsymbol{x}|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta}), \tag{5}$$

where the evaluation of the prior $p(\boldsymbol{\vartheta})$ is immediate while the evaluation of likelihood-to-evidence ratio $\hat{r}(\boldsymbol{x}|\boldsymbol{\vartheta})$ only requires a single forward pass through the neural network – *an operation which is fast and easily parallelizable in comparison to sampling algorithms.*

This amortization procedure is generic and can be applied to any Bayesian posterior inference task. In particular, it also applies for the fast evaluation of marginal posterior distributions defined over a subset of parameters of interest, such as the posterior over the component masses $m_1 - m_2$, the posterior over the mass ratio and the effective spin $q - \chi_{\text{eff}}$, or the posterior over right ascension and declination $\alpha - \delta$. In this case, the corresponding marginal likelihood functions are typically not available analytically and the posterior sampling procedure must be run over the entire parameter space for the only purpose of marginalizing out the other parameters.

4 Network architecture and training

The neural network used to model the likelihood-to-evidence ratio and to demonstrate our inference method on GW signals is illustrated in Figure 1. The Hanford and Livingston detectors input strains $\boldsymbol{x}$ are processed by a convolutional trunk made of 13 layers of dilated convolutions. The output log-ratio is computed by a 3-layer fully connected network ending with no activation function and taking as input both the convolutional feature map and the parameters $\boldsymbol{\vartheta}$. At training time, the class $y = 1$ probability $p(y = 1|\boldsymbol{x}, \boldsymbol{\vartheta})$ is computed as the sigmoid activation $\sigma(\log r(\boldsymbol{x}|\boldsymbol{\vartheta}))$ of the network's output.

To train and evaluate our neural network model, we consider simulated signals of binary black hole mergers using `IMRPhenomPv2` as approximant (including the effects of spin precession), and parameters drawn from the prior distribution $p(\boldsymbol{\vartheta})$ summarized in Table 2. To ease comparison with standard analysis results, the prior is chosen identical to the analysis of GW150914 [3] and the companion analysis of PYCBC [7]. Our data generation process follows the procedure of [15] in order to produce realistic synthetic training data. In short, background noise is emulated using a Gaussian noise model with a PSD identical to the one used in [2]. Simulated signals are injected into the noise and the result is then whitened and high-passed at 20 Hz to remove simulation artefacts. Finally, the generated strains are cropped to a length of $4s$ in such a way that the maximum of the signal always ends up at a random location in a fixed interval of $0.2s$ within the sample. Using a sampling rate of 2048 Hz, the resulting observable $\boldsymbol{x}$ for each realization is an array of size 8192×2, where the channels correspond to the strains from each of the two interferometers. Using this procedure, we generated 10^6 observables for training, 2×10^5 for validation and 2×10^5 for test.
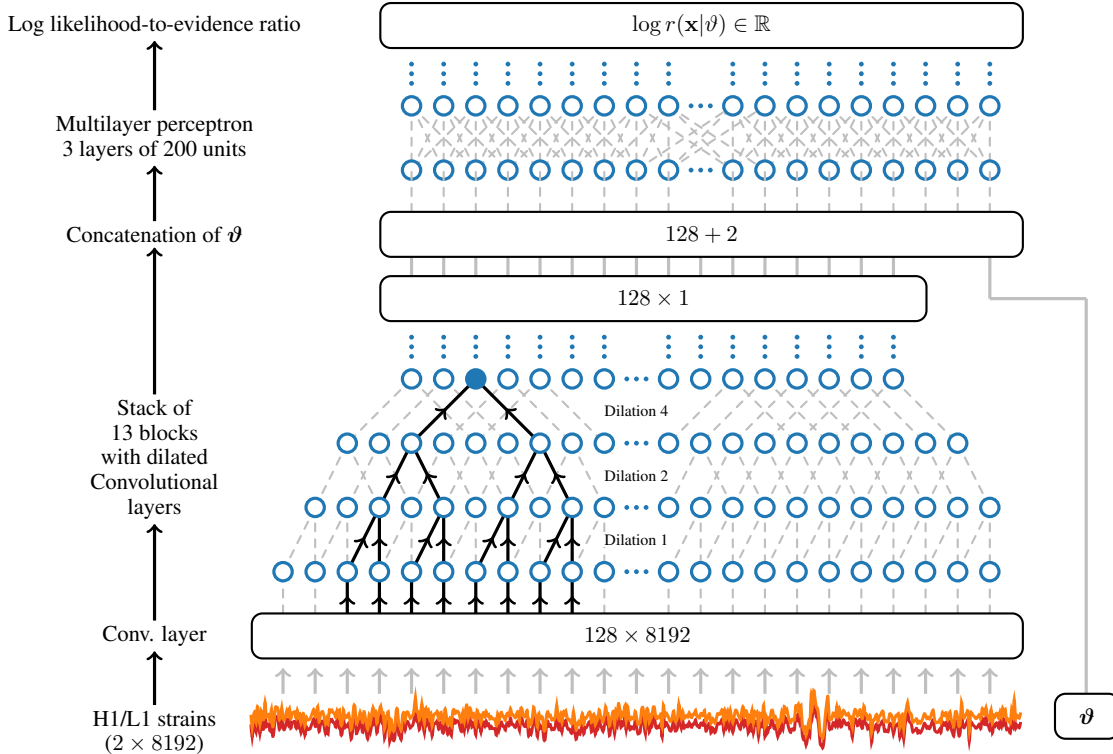
Figure 1: The architecture of the neural ratio estimator, adapted from [15]. A feature map of both the Hanford's and Livingston's strains is produced by a CNN composed of a 1D convolutional layer (128 kernels of size 1) followed by 13 dilated convolutional layers (128 kernels of size 2). Finally, the convolutional feature map (128 channels of size 1) is concatenated with the parameters ϑ and fed to a 3-layer fully connected network for approximating the log likelihood-to-evidence ratio.

5   Results

To evaluate the performance of our method, we perform inference on simulated gravitational waves. Figure 2 showcases test GWs and the resulting approximate posteriors. The first row compares the credible intervals obtained with our method and with MCMC on a signal similar to GW150914. Both MCMC and our method produce consistent results but *our method takes less than a minute to run while MCMC ran for more than a day!* Since performing an MCMC run is computationally expensive, the following rows only compare against the nominal parameter values that were used to generate the signals. We observe that our method is often able to infer the parameters with high precision and to produce narrow credible intervals. Sometimes, our method produces wide intervals.

We look at the coverage on 1000 simulated signals to evaluate the reliability of the estimated credible intervals. If the model estimates correctly the credible intervals, then the empirical coverage probability should be close to the nominal coverage probability. For instance, the estimated 50%-credible intervals for pairs $x, \vartheta \sim p(x, \vartheta)$ should contain the nominal value ϑ approximately 50% of the time. As reported in Table 5, the empirical coverage probability is usually slightly higher than expected. This shows that the model is slightly under-confident in its predictions and hence produces credible intervals than are slightly larger than expected. This under-confidence of the model is however moderated and does not preclude the contours to be useful in practice since predictions are conservative.

6   Conclusions

In this paper, we provide a proof of concept that demonstrates that neural simulation-based inference, and more specifically likelihood-to-evidence ratio estimation, can be used to speed up the analysis

	$m_{max}^{det}, m_{min}^{det}$	d_L, θ_{JN}	χ_{eff}, q	α, δ
50% credible interval coverage	51.8%	50.7%	57.5%	54.2%
90% credible interval coverage	90.2%	89.8%	93.3%	92.8%

Table 1: Empirical coverage probability of the estimated credible intervals.

of GWs by up to three orders of magnitude while producing good posterior parameter distributions. Although these preliminary results are promising, we note that there are still obstacles to overcome before deploying neural simulation-based inference in official GW analysis pipelines. Possible improvements include accounting for the measured noise PSD, as well as further exploring the preprocessing of the data. Further assessments of the statistical validity of the estimated posteriors would also be needed before making any reliable scientific claims. Recently, complementary deep learning approaches such as normalizing flows and conditional variational auto-encoders have also been applied to the same problem. A detailed comparison between all those methods is required to further advance the field.

Broader impact

Neural simulation-based inference has the potential to speed up scientific discoveries by allowing the fast analysis of gravitational waves. It also unlocks multi-messenger astronomy as the fast inference of the binary black-holes merger's sky position is required to measure electromagnetic and/or astroparticle counterparts to the GW signal. Malicious usage of our research is hard to imagine. However, failure modes of our inference engine could lead to erroneous scientific claims, and thus should be carefully validated.

Acknowledgments

References

[1] J. Aasi et al. Advanced LIGO. *Class. Quant. Grav.*, 32:074001, 2015.

[2] B. Abbott, R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. Adhikari, V. Adya, C. Affeldt, et al. Gwtc-1: a gravitational-wave transient catalog of compact binary mergers observed by ligo and virgo during the first and second observing runs. *Physical Review X*, 9(3):031040, 2019.

[3] B. Abbott et al. Properties of the Binary Black Hole Merger GW150914. *Phys. Rev. Lett.*, 116(24):241102, 2016.

[4] B. Abbott et al. Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Rev. Rel.*, 21(1):3, 2018.

[5] B. P. Abbott et al. A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals. *Class. Quant. Grav.*, 37(5):055002, 2020.

[6] F. Acernese et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. *Class. Quant. Grav.*, 32(2):024001, 2015.

[7] C. Biwer, C. D. Capano, S. De, M. Cabero, D. A. Brown, A. H. Nitz, and V. Raymond. PyCBC Inference: A Python-based parameter estimation toolkit for compact binary coalescence signals. *Publ. Astron. Soc. Pac.*, 131(996):024503, 2019.

[8] J. Brehmer, S. Mishra-Sharma, J. Hermans, G. Louppe, and K. Cranmer. Mining for Dark Matter Substructure: Inferring Subhalo Population Properties from Strong Lenses with Machine Learning. ApJ, 886(1):49, Nov 2019.

[9] A. J. Chua and M. Vallisneri. Learning Bayesian posteriors with neural networks for gravitational-wave inference. *Phys. Rev. Lett.*, 124(4):041102, 2020.

[10] K. Cranmer, J. Brehmer, and G. Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 2020.

[11] K. Cranmer, J. Pavez, and G. Louppe. Approximating Likelihood Ratios with Calibrated Discriminative Classifiers. *arXiv preprint arXiv:1506.02169*, 2015.

[12] E. Cuoco, J. Powell, M. Cavaglià, K. Ackley, M. Bejger, C. Chatterjee, M. Coughlin, S. Coughlin, P. Easter, R. Essick, et al. Enhancing gravitational-wave science with machine learning. *arXiv preprint arXiv:2005.03745*, 2020.

[13] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith. Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy. *arXiv preprint arXiv:1909.06296*, 2019.

[14] H. Gabbard, M. Williams, F. Hayes, and C. Messenger. Matching matched filtering with deep networks for gravitational-wave astronomy. *Phys. Rev. Lett.*, 120(14):141103, 2018.

[15] T. D. Gebhard, N. Kilbertus, I. Harry, and B. Schölkopf. Convolutional neural networks: a magic bullet for gravitational-wave detection? *Phys. Rev. D*, 100(6):063015, 2019.

[16] D. George and E. Huerta. Deep Learning for Real-time Gravitational Wave Detection and Parameter Estimation: Results with Advanced LIGO Data. *Phys. Lett. B*, 778:64–70, 2018.

[17] D. George and E. Huerta. Deep Neural Networks to Enable Real-time Multimessenger Astrophysics. *Phys. Rev. D*, 97(4):044039, 2018.

[18] S. R. Green and J. Gair. Complete parameter inference for gw150914 using deep learning. *arXiv preprint arXiv:2008.03312*, 2020.

[19] S. R. Green, C. Simpson, and J. Gair. Gravitational-wave parameter estimation with autoregressive neural network flows. *arXiv preprint arXiv:2002.07656*, 2020.

[20] M. Hannam, P. Schmidt, A. Bohé, L. Haegel, S. Husa, F. Ohme, G. Pratten, and M. Pürrer. Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms. *Phys. Rev. Lett.*, 113(15):151101, 2014.

[21] J. Hermans, V. Begy, and G. Louppe. Likelihood-free mcmc with amortized approximate ratio estimators. In *International Conference on Machine Learning*, 2020.

[22] X. Li, W. Yu, and X. Fan. A method of detecting gravitational wave based on time-frequency analysis and convolutional neural networks. *arXiv preprint arXiv:1712.00356*, 2017.

[23] P. S. Shawhan, P. R. Brady, A. Brazier, S. Cenko, M. Jurić, and E. Katsavounidis. Data Analysis Challenges for Multi-Messenger Astrophysics. *ASP Conf. Ser.*, 523:705, 2019.

[24] H. Shen, E. Huerta, Z. Zhao, E. Jennings, and H. Sharma. Deterministic and bayesian neural networks for low-latency gravitational wave parameter estimation of binary black hole mergers. *arXiv preprint arXiv:1903.01998*, 2019.

[25] J. Veitch et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Phys. Rev.*, D91(4):042003, 2015.

A Likelihood function and prior distribution

Our likelihood function (which is only defined up to an overall constant factor) is given by

$$p(d(t)|\boldsymbol{\vartheta}, h(t, \boldsymbol{\vartheta})) =$$

$$\exp\left\{-\frac{1}{2}\sum_{k=1}^{N}\left\langle\tilde{d}_k(f) - \tilde{h}_k(f, \boldsymbol{\vartheta})\Big|\tilde{d}_k(f) - \tilde{h}_k(f, \boldsymbol{\vartheta})\right\rangle\right\}, \tag{6}$$

where $\langle a|b\rangle$ denotes the noise-weighted inner product, defined as

$$\left\langle\tilde{a}_k(f)\Big|\tilde{b}_k(f)\right\rangle = 4\Re\int_0^\infty \frac{\tilde{a}_k(f)\tilde{b}_k(f)}{S_k(f)}\,df\,, \tag{7}$$

$d(t)$ are the GW detector data time series, N is the number of detectors, $\tilde{d}_k(f)$ is the Fourier transform of the data from the k-th detector, $S_k(f)$ is the noise power spectral density (PSD) of the k-th detector, and $\tilde{h}_k(f, \boldsymbol{\vartheta})$ is the frequency domain model waveform projected onto the k-th detector. Specifically, this is the sum of the $+$ and $\times$ GW polarizations weighted by the antenna response factors $F^{+,\times}$, which depend on the extrinsic parameters relating the source position and orientation to the detectors, i.e.

$$\tilde{h}_k(f, \boldsymbol{\vartheta}) = F_k^+ \tilde{h}^+(f, \boldsymbol{\vartheta}) + F_k^\times \tilde{h}^\times(f, \boldsymbol{\vartheta}). \tag{8}$$

Table 2 summarizes the signal model parameters and their respective priors, as used in the training of the inference network. We adopt the same prior as in [3, 7], to which we refer for further details.

Parameter	Symbol	Unit	Prior	Minimum	Maximum
First/Second BH mass	$m_{1,2}$	$M_\odot$	uniform	10	80
Spin vectors	$\boldsymbol{S}_{1,2}$	$-$	$-$	$-$	$-$
$-$ magnitudes		$-$	uniform	0	0.99
$-$ polar angles		rad	} uniform solid	0	π
$-$ azimuthal angles		rad		0	2π
Luminosity distance	d_L	Mpc	uniform radius	10	1000
Right ascension	α	rad	} uniform sky	0	2π
Declination	δ	rad		$-\pi/2$	$\pi/2$
Binary inclination angle	$\theta_{\rm JN}$	rad	sin	0	π
Coalescence phase angle	Φ	rad	uniform	0	2π
Polarization angle	Ψ	rad	uniform	0	2π
Time shift	$t_{\rm shift}$	s	uniform	-0.1	0.1

Table 2: Prior distribution $p(\boldsymbol{\vartheta})$ on the signal model parameters.
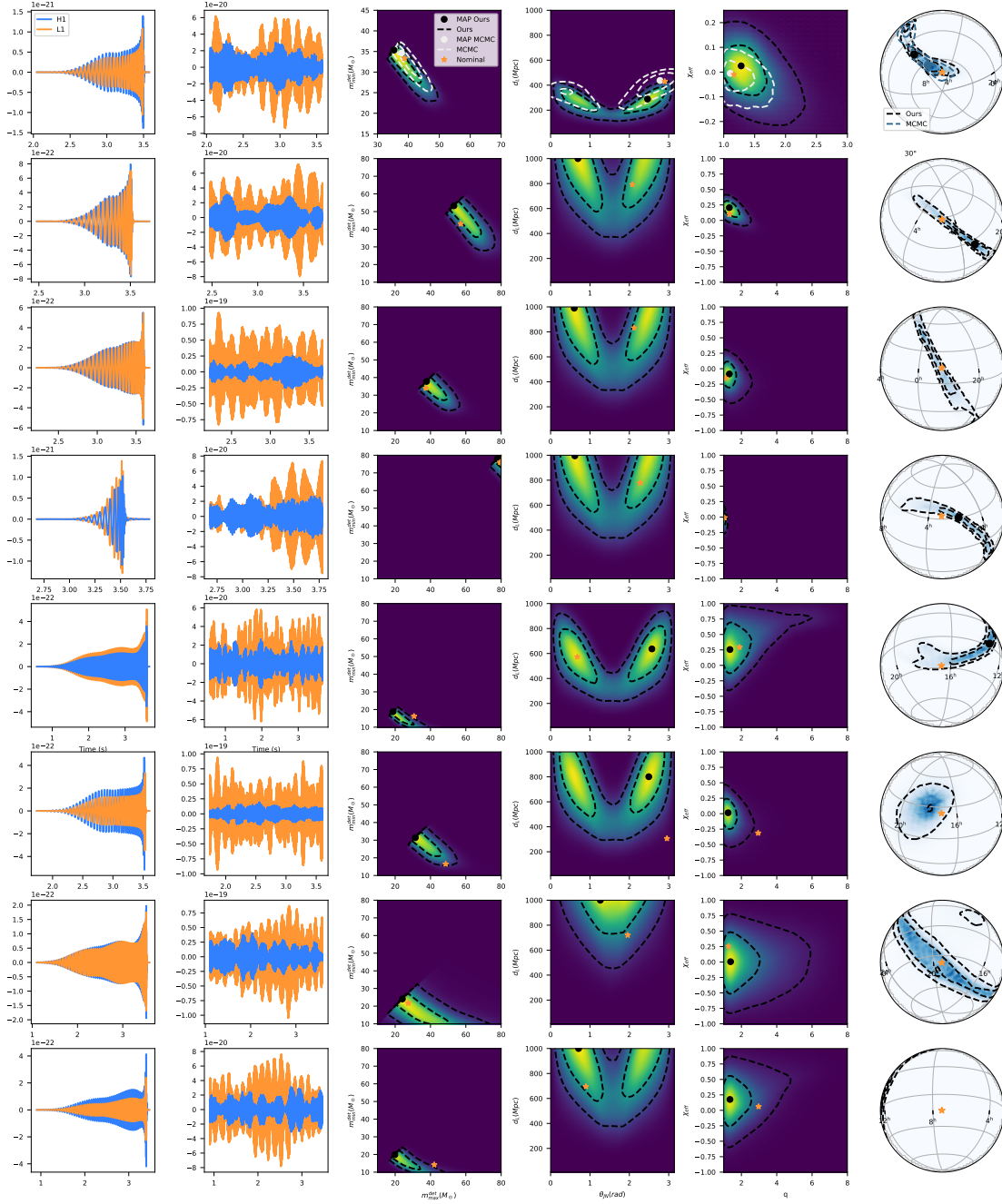
B Mock gallery

Figure 2: Results obtained with our method on simulated gravitational waves. The first column shows the gravitational wave without noise. The second column shows the same gravitational wave with noise added. The last 4 columns show the results obtained when performing posterior parameter inference from the noisy gravitational wave. For each sample, the 50% and 90% credible intervals as well as the Maximum a posteriori estimates are shown. Those are compared to the nominal parameter values used to generate the data. In the first row, our results are compared to results obtained using MCMC.