Identification and characterization of distorted gravitational waves by lensing using deep learning

Juno C. L. Chan,[a] Lorena Magaña Zertuche,[b] Jose María Ezquiaga, Rico K. L. Lo, Luka Vujeva, and Joey Bowman
Center of Gravity, Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, Denmark
(Dated: November 11, 2025)

Gravitational waves (GWs) can be distorted by intervening mass distributions while propagating, leading to frequency-dependent modulations that imprint a distinct signature on the observed waveforms. Bayesian inference for GW lensing with conventional sampling methods is costly, and the problem is exacerbated by the rapidly growing GW catalog. Moreover, assessing the statistical significance of lensed candidates requires thousands, if not millions, of simulations to estimate the background from noise fluctuations and waveform systematics, which is infeasible with standard samplers. We present a novel method, `DINGO-lensing`, for performing inference on lensed GWs, extending the neural posterior estimation framework `DINGO`. By comparing our results with those using conventional samplers, we show that the compute time of parameter estimation of lensed GWs can be reduced from weeks to seconds, while preserving accuracy both in the posterior distributions and the evidence ratios. We train our neural networks with LIGO detector noise at design sensitivity and a lens model that accommodates two overlapping chirps with opposite parity. We show that the lensing parameters are recovered with millisecond precision for the time delays. We also demonstrate that our network can identify signals diffracted by point masses, highlighting its flexibility for searches. By simulating thousands of lensed and nonlensed events, we determine how the detectability changes with different source properties. `DINGO-lensing` provides a scalable and efficient avenue for identifying and characterizing gravitationally lensed GW events in the upcoming observing runs.

arXiv:2511.07186v1 [gr-qc] 10 Nov 2025

I. INTRODUCTION

The strongest emission of gravitational waves (GWs) in the Universe comes from compact binary coalescences (CBCs), such as binary neutron stars, neutron star-black hole systems, and binary black holes (BBHs). The signals coming from these highly energetic events are detected by the LIGO-Virgo-KAGRA (LVK) detector network [1–3], which has already brought us many confident BBH GW detections [4]. By comparing the observed signals to those coming from numerical relativity or a robust model, we are able to estimate the intrinsic and extrinsic parameters of such binary systems. In order to do this accurately, we rely on cutting-edge models that capture several physical effects, such as higher harmonic modes, eccentricity, and precession [5–16].

GWs propagate over cosmological distances and are not impervious to intervening matter. Instead, their path is affected by nearby gravitational potentials (acting as lenses), such as those of stars, black holes, galaxies, and clusters of galaxies, leading to gravitational lensing [17, 18]. In strong gravitational lensing, multiple signals, or images, of the source are produced each with a distinct amplitude, phase shift, and time delay [19]. The arrival time difference between these images can be very small for smaller lens masses (or specific source-lens-observer configurations) [19–23], making it challenging to look for signals coming from the same source. Additionally, these small lens masses may produce overlapping images which causes a beating pattern in the inspiral portion of the waveform resembling precession or the effect of two different binary signals overlapping [24, 25]. In the full wave-optics regime, it is also possible to have a single signal with frequency-dependent modulations due to diffraction effects. Detecting such signal can be challenging in matched-filtering searches [26], but it can answer key questions in the small lens mass regime regarding small compact lens populations [27–41], the halo mass function [42–45], the structures of different lenses [20–23, 46–51], the cosmic expansion rate [52–54], and the astrophysical environments of GW sources [55–57], as well as provide an alternate way to test for deviations of general relativity [58–62].

Confirming the lensed nature of a candidate event requires scrutiny, even in small effects such as noise fluctuations and waveform systematics [63]. Making a confident detection claim requires the simulation and analysis of millions of events. Carefully conducted searches for lensing signatures in GWs in the first three observing runs found no convincing evidence for GW lensing [64–69]. However, the fourth observing run brought a lensed candidate, GW231123 [70], which we investigate in detail in a companion paper [71].

The high computational cost of performing Bayesian inference on distorted GW signals with conventional sampling algorithms, such as Markov Chain Monte Carlo and nested sampling [72, 73], poses a major limitation for lensing studies. Not only does it prevent us from keeping pace with the rapidly growing number of detections expected from an expanding detector horizon, but it also makes background estimation for lensing searches prohibitively expensive, as millions of nonlensed parameter

[a] chun.lung.chan@nbi.ku.dk
[b] lorena.zertuche@nbi.ku.dk

estimation (PE) runs are required to quantify the lensing false-alarm rate. To address these challenges, several alternatives that bypass conventional Bayesian inference have been proposed [74, 75]. While these methods offer significant speedups, they often lack the full statistical interpretability and robustness of standard Bayesian inference.

Deep-learning techniques have emerged as a powerful solution for performing fast and accurate PE of GW signals [73]. Unlike common Bayesian sampling methods — which require repeated waveform evaluations and are therefore computationally expensive — neural networks can learn the complex mapping between GW strain data and source parameters directly from simulated training sets. Once trained, these models can generate full posteriors within seconds, achieving orders-of-magnitude speedups while maintaining precision comparable to traditional PE pipelines [76–87]. Recent studies have begun extending deep-learning techniques to the lensed GW regime, exploring the use of convolutional-, recurrent-, and transformer-based architectures for rapid identification of lensing signatures [88–90] and inferring lens parameters [75, 91].

Among these approaches, the `DINGO` framework [87, 92–99] represents a state-of-the-art implementation of simulation-based inference using neural posterior estimation (NPE). `DINGO` combines an embedding network with normalizing flows to produce full Bayesian posteriors conditioned on GW data, enabling rapid and high-fidelity inference across high-dimensional parameter spaces. Simulation-based inference provides a solution by combining computational efficiency and statistical interpretability. Instead of relying on explicit likelihood evaluations, simulation-based inference employs neural density estimators trained on simulated data to learn the complex mapping between observations and parameters [73, 100].

Building on top of the simulation-based inference algorithm, `DINGO`, we present a new lensing inference pipeline `DINGO-lensing`[1] to perform NPE. In particular, we employ the geometric optics approximation in the wave optics regime in our lensed waveform generation. By directly learning the posterior distribution from simulated data, `DINGO-lensing` circumvents the computational bottlenecks of a traditional sampler while maintaining the statistical rigour of conventional inference. This approach establishes a scalable and robust pathway for identifying and characterizing lensed GW events in the era of high-cadence detections.

This paper is structured as follows: Section II describes the lensed waveform model, network architecture, and training setup. Section III presents the results and validation of our trained network, including comparisons with traditional Bayesian inference, systematics, and an assessment of detectability. Finally, Section IV summarizes our findings and discusses the implications for future applications of simulation-based inference in lensing analyses.

[1] https://github.com/dingo-lensing/dingo-lensing

II. `DINGO-LENSING`

In this work, we perform NPE training with the simulation-based inference framework `DINGO`. We use the fifth observing run (O5) noise amplitude spectral density (ASD) [101, 102], assuming Gaussian and stationary noise. Throughout the study, we utilize the inspiral-merger-ringdown phenomenological model with higher modes and precession, `IMRPhenomXPHM` [9] implemented in `LALSimulation` [103, 104], with a frequency range of $[20, 1024]$ Hz and a waveform duration of 8 s. We have modified the network to allow for the use of numerical relativity surrogate waveforms during training and parameter inference, e.g., `NRSur7dq4` that spans a frequency range of [20, 512] Hz, which we use in the companion paper [71]. The intrinsic parameters consist of the component masses in the detector frame m_i $(i = 1, 2)$, the spin magnitudes a_i, and the spin angles $(\phi_i, \phi_{12}, \phi_{JL})$. Other parameters included are the inclination angle θ_{JN}, right ascension (RA) and declination (DEC), polarization angle ψ, luminosity distance $d_{\rm L}$, the time of the peak amplitude at geocenter $t_{\rm c}$, and the phase at a reference frequency $\phi_{\rm ref}$ of 20Hz.

The following subsections present the lensed waveform model adopted in constructing the training dataset and in the subsequent analyses, along with a description of the training setup.

A. Lensed Waveform Model

With the weak-gravity and thin-lens approximation, the gravitational lensing of a GW signal $h(t)$ can be solved efficiently in the frequency domain [19]. The lensed $\tilde{h}_{\rm L}(f)$ and unlensed $\tilde{h}_{\rm U}(f)$ frequency-dependent GW signals are related by the amplification factor $F(f)$ as

$$\tilde{h}_{\rm L}(f) = F(f)\tilde{h}_{\rm U}(f)\,. \tag{1}$$

The amplification factor is obtained by solving the diffraction integral,

$$F(f, \boldsymbol{\theta}_{\rm S}) = \frac{\tau_D f}{i} \int {\rm d}^2\boldsymbol{\theta}\, e^{2\pi i f t_{\rm d}(\boldsymbol{\theta}, \boldsymbol{\theta}_{\rm S})}\,, \tag{2}$$

which integrates over the image plane positions $\vec{\theta}$. Here, τ_D is the characteristic timescale associated to the distances between source-lens-observer, $\boldsymbol{\theta}_{\rm S}$ is the nonlensed source angular position, and $t_{\rm d}(\boldsymbol{\theta}, \boldsymbol{\theta}_{\rm S})$ is the time delay between the source and observer, often called the *time-delay surface*.

Using the stationary phase approximation (SPA), the diffraction integral simplifies to a sum over multiple copies of the original signal. Each of these copies can be magnified, time delayed and phase shifted. If the time delay between any of the repeated copies is less than the signal's duration, then there are interference effects that distort the signal. In this work we focus on an amplification factor that depends on the time delay Δt and relative magnification $\mu_{\rm rel}$ between two images with a phase difference of $\pi/2$, i.e.,

$$F(f>0) = 1 + \sqrt{\mu_{\rm rel}}\, e^{i(2\pi f \Delta t - \pi/2)} \,. \tag{3}$$

This is a good approximation for systems in which two images are interfering with each other and has been shown to hold for several types of singularities, including point, fold, and (some) cusp cases (see [22, 105] where further details on its validity can be found). Operationally, we implement this lens model coded in `modwaveforms`.[2]

Image properties are determined by their locations on the time delay surface. That is, images that form on minima of the time delay surface are of Type I, those that form on saddle points are of Type II with a $\pi/2$ phase shift, and those on maxima are of Type III with a π phase shift. For example, a point-mass lens always produces two images of Type I and II. For the point, fold, and many configurations near cusp singularities, the brightest image (of Type I) arrives before the second brightest image (of Type II). This allows us to define our relative magnification, or the ratio of the second brightest to the brightest image $\mu_{\rm rel} \equiv \mu_2/\mu_1$, as being $\mu_{\rm rel} \in [0,1]$ and the time delay as $\Delta t > 0$. This will always result in a $\pi/2$ phase difference between the two lensed images. This assumption removes a possible degeneracy where images of the same type interfering with each other can lead to $(\mu_{\rm rel}, \Delta t) = (0,0)$ or $(\mu_{\rm rel}, \Delta t) = (1,0)$, which is broken in our case due to the phase difference between the two images. Therefore, within our framework we define a "lensed image" as an event in which our analysis finds $\Delta t > 0$, implying $\mu_{\rm rel} < 1$. We define a "nonlensed event" as an event which our analysis finds a relative magnification factor of $\mu_{\rm rel} = 0$. Exploring lensing configurations where one would find overlapping images of the same type, as well as images with $\mu_{\rm rel} > 1$ such as the ones created by lenses containing substructure [21, 23], is left to future work.

B. Training Setup

We perform GW parameter estimation using `DINGO-lensing`, building on top of the inference pipeline `DINGO` [92–97]. We use NPE to model the Bayesian conditional distribution probability $q_\phi(\boldsymbol{\theta}|d)$ as a surrogate for the true posterior probability $p(\boldsymbol{\theta}|d)$, given by

$$p(\boldsymbol{\theta}|d) = \frac{\mathcal{L}(d|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\mathcal{Z}(d)}, \tag{4}$$

where ϕ are the training parameters in the network. Training is carried out on simulated, labeled pairs $(\boldsymbol{\theta}, d)$, where the parameters $\boldsymbol{\theta}$ are drawn from the prior $\pi(\boldsymbol{\theta})$, and the corresponding detector strain d is sampled from the likelihood $\mathcal{L}(d|\boldsymbol{\theta})$. In practice, this involves injecting a waveform $h(\boldsymbol{\theta}_i)$ into additive Gaussian noise n_i, assuming the noise is stationary: $d_i = h(\boldsymbol{\theta}_i) + n_i$.

The network architecture follows the design in [95], with all hyperparameters kept unchanged. In short, `DINGO-lensing` employs a dual-network structure: an embedding network, whose first layer reduces the raw strain data to a 128-dimensional representation based on singular value decomposition (SVD), followed by a fully connected network that maps this representation into an approximation of the posterior. Our architecture comprises $\sim 10^8$ learnable parameters and $\sim 10^7$ fixed parameters.

We generate training datasets consisting of 5×10^6 simulated waveforms, with parameters drawn from the priors listed in Table I. The waveforms are stored in an SVD representation, and 2% of the dataset is reserved for testing.

We adopt the two-stage training scheme with the hyperparameters, where the initial training stage (Stage 0) freezes the initial SVD layer and is unfreezed in the fine-tuning stage (Stage 1). The training of both the lensed and nonlensed NPE models is performed in two stages. During Stage 0, the model is trained for 300 epochs using the Adam optimizer [106] with a learning rate of 7×10^{-5} and a cosine learning-rate scheduler [107]. In Stage 1, all layers are unfrozen for fine-tuning over an additional 150 epochs, with the learning rate reduced to 1×10^{-5}. Both

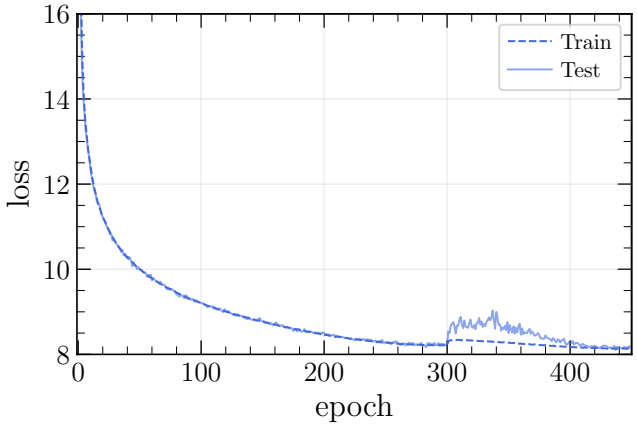

Figure 1. The loss as a function of the number of training epochs of the lensed network, shown for both the training set (dashed) and testing set (solid).

[2] https://github.com/ezquiaga/modwaveforms

Parameter	Unit	Prior	Injection
$\mathcal{M}_c$	$M_\odot$	$\mathcal{U}(10, 100)$	49.2
q	$M_\odot$	$\mathcal{U}(0.125, 1.0)$	0.79
m_1	$M_\odot$	$\mathcal{C}(5, 150)$	63.6
m_2	$M_\odot$	$\mathcal{C}(5, 150)$	50.4
a_1		$\mathcal{U}(0, 0.99)$	0.28
a_2		$\mathcal{U}(0, 0.99)$	0.45
ϕ_{12}		$\mathcal{U}(0, 2\pi)$	3.59
ϕ_1		$\sin(0, \pi)$	1.75
ϕ_2		$\sin(0, \pi)$	0.49
ϕ_{JL}		$\mathcal{U}(0, 2\pi)$	5.38
θ_{JN}		$\sin(0, \pi)$	2.83
$\phi_{\rm ref}$		$\mathcal{U}(0, 2\pi)$	1.73
RA		$\mathcal{U}(0, 2\pi)$	3.01
DEC		$\cos(-\pi/2, \pi/2)$	-0.13
ψ		$\mathcal{U}(0, \pi)$	2.63
t_c	s	$\mathcal{U}(-0.1, 0.1)$	-0.0565
d_L	Gpc	$\mathcal{U}(1, 15)$	9.84
Δt	s	$\mathcal{U}(0, 0.1)$	0.031
$\mu_{\rm rel}$		$\mathcal{U}(0, 1)$	0.79

Table I. The first three columns show the prior distributions corresponding to a given parameter used in the training dataset for NPE, where $\mathcal{U}$ denotes a uniform distribution and $\mathcal{C}$ denotes a parameter is constrained to be in such range. The last column has the injected values for comparing the accuracy of `DINGO-lensing` and `bilby` in Figs. 3 and 10.

stages employ a batch size of 4096 and temperatures of 300 and 150, respectively. This two-stage procedure stabilizes early optimization and enables the network to refine the embedding representation for improved posterior accuracy.

During the training process, a neural network evaluates its performance by computing a *loss function*, which measures the discrepancy between the predicted and target probability distributions. The loss function corresponds to the negative log-likelihood of the data, under the forward Kullback–Leibler divergence model, between the true posterior $p(\boldsymbol{\theta}|d)$ and the neural approximation $q_\phi(\boldsymbol{\theta}|d)$ [93].

Minimizing this loss encourages the network to produce posteriors that closely match those obtained from exact Bayesian inference. The optimization proceeds iteratively over epochs by adjusting the network parameters ϕ, typically via stochastic gradient descent, until convergence, as exemplified in Figure 1.

C. Importance Sampling

To obtain accurate posterior estimates without retraining the neural network for each new likelihood configuration, `DINGO` employs importance sampling as a correction step following the NPE stage [97]. Given a trained model that approximates the posterior distribution $q_\phi(\boldsymbol{\theta}|d)$ over source and lensing parameters $\boldsymbol{\theta}$ conditioned on data d, we reweight samples from this approximate posterior to the true target posterior under a desired likelihood $p(d|\boldsymbol{\theta})$ and prior $p(\boldsymbol{\theta})$.

The corrected posterior is obtained via the standard importance reweighting relation:

$$p(\boldsymbol{\theta}|d) \propto \frac{p(d|\boldsymbol{\theta})p(\boldsymbol{\theta})}{q_\phi(\boldsymbol{\theta}|d)}. \tag{5}$$

Given a set of N samples $\{\boldsymbol{\theta}_i\}_{i=1}^N$ drawn from $q_\phi(\boldsymbol{\theta}|d)$, the corresponding importance weights are computed as

$$w_i = \frac{p(d|\boldsymbol{\theta}_i)\, p(\boldsymbol{\theta}_i)}{q_\phi(\boldsymbol{\theta}_i|d)}, \qquad \tilde{w}_i = \frac{w_i}{\sum_{j=1}^N w_j}, \tag{6}$$

where $\tilde{w}_i$ are the normalized weights satisfying $\sum_i \tilde{w}_i = 1$. Expectations of any quantity $f(\boldsymbol{\theta})$ under the target posterior are then estimated as

$$\mathbb{E}_p[f(\boldsymbol{\theta})] \approx \sum_{i=1}^N \tilde{w}_i\, f(\boldsymbol{\theta}_i). \tag{7}$$

In practice, the importance weights are computed using the likelihood function implemented in `bilby`, ensuring that the reweighted samples are consistent with the full noise model and data segment used for parameter estimation. This step allows the neural posterior to flexibly adapt to small mismatches between the simulation-based training distribution and the actual data likelihood, without retraining the network. The efficiency of this procedure is monitored via the sampling efficiency,

$$\epsilon = \frac{1}{N} \frac{\left(\sum_i \tilde{w}_i\right)^2}{\sum_i \tilde{w}_i^2}, \tag{8}$$

which quantifies the effective number of posterior samples contributing to the reweighted estimate. In our analyses, ϵ typically remains above a few percent, indicating that the learned posterior provides a sufficiently close approximation to the true one for reliable reweighting.

D. Lensing Bayes Factor

To quantify the support for the lensed hypothesis over the not lensed hypothesis, we compute the Bayes factor

$$\mathcal{B}_{\rm lens} = \frac{p(d|\mathcal{H}_{\rm L})}{p(d|\mathcal{H}_{\rm NL})}, \tag{9}$$

where $\mathcal{H}_{\rm L}$ and $\mathcal{H}_{\rm NL}$ denote the lensed and not lensed hypotheses, respectively.

The Bayesian evidence under each hypothesis is approximated via importance sampling. Specifically, samples are drawn from the neural posterior estimator and reweighted according to the true likelihood-to-network ratio, following the approach of [97]. This procedure allows for an efficient and unbiased estimate of the marginal likelihood without requiring explicit integration over the full parameter space. We train separate non-lensed and lensed models to perform posterior estimation under each hypothesis with the training setup and validation described in Appendix A.

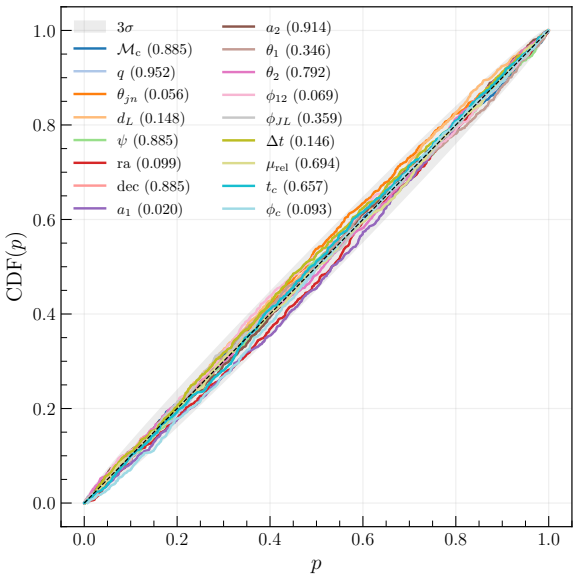

Figure 2. P-P plot for the lensed neural posterior estimation model using 10^3 lensed injections and no importance sampling. For each injection, we generate a posterior with our model and compute the percentile value of the injected parameter. Each colored line corresponds to the cumulative distribution function (CDF) of the corresponding parameters (see label). The Kolmogorov-Smirnov test p-values are given in the legend.

III. RESULTS

In this section, we present the training results of our lensed network and summarize its performance through a series of model validation tests. We further compare representative corner plots with `bilby` and demonstrate that `DINGO-lensing` can effectively recover point-lens injections.

A. Training performance

Figure 1 shows the evolution of the loss as a function of training epoch. The training loss (dashed) and test loss (solid) remain closely aligned throughout, indicating the absence of overfitting. The smooth behavior of both curves further demonstrates the stability of the training process. After approximately 400 epochs, the loss function reaches a plateau, suggesting that the model has ceased to learn significant new information from the data. A noticeable discontinuity around epoch 300 corresponds to unfreezing the additional layers in stage 1. We terminate the training at epoch 450, as no further improvement in accuracy is observed. The training takes ~ 10 days with an NVIDIA A100 PCIe GPU.

B. Network validation

We perform inference on 10^3 lensed injections generated from our trained model and draw 10^4 samples for each injections from the lensed network. We then reconstruct the reference phase [97]. Figure 2 displays the corresponding probability–probability (P–P) plot constructed directly from the posteriors obtained by the lensed network without applying importance sampling. The diagonal line represents the ideal case in which $x\%$ of the true parameters lie within the $x\%$ credible intervals of the inferred posteriors. The colored curves for individual parameters show excellent agreement with this diagonal, demonstrating that the network produces well-calibrated posteriors. This proper calibration confirms that the sampling procedure functions as intended, a result further supported by the Kolmogorov–Smirnov test p-values shown in the legend. Minor deviations for the lensing parameters (Δt and $\mu_{\rm rel}$) remain within expected statistical fluctuations (3σ), indicating robust performance of the model even in the presence of lensing effects.

In addition, we look at the distribution of sampling efficiencies ϵ after importance sampling. We find that they span a wide range of values between $\sim 10^{-4}$ and ~ 0.1 with no particular correlation with the binary and lens parameters. Moreover, we have tested the effects of sampling efficiency on the Bayes factor computation and find good agreement between `DINGO-lensing` and `bilby` for $\epsilon > 10^{-4}$. For more details see appendix B.

C. Comparison with traditional parameter estimation methods

Figure 3 shows a corner plot comparing the posterior distributions obtained using our `DINGO-lensing` model (solid blue curves) with the reference `bilby` inference pipeline (dashed gray curves) for a representative lensed injection with a network SNR = 16.18 at O5 sensitivity. The source parameters and priors used in the `bilby` analysis are tabulated in Table I.

The one- and two-dimensional marginalized posteriors show excellent agreement between `DINGO-lensing` and `bilby` across both intrinsic (e.g., $\mathcal{M}_c$, q, $a_{1,2}$) and extrinsic (e.g., d_L, θ_{JN}, ψ) parameters, demonstrating that the neural posterior estimator accurately reproduces the full Bayesian inference results at a fraction of the computational cost. The true parameters (solid gray lines) are also well recovered within the inferred 68% credible regions (shaded regions in dark blue; while 95% credible regions are shared in light blue) with a sampling efficiency of 2.3%. The Bayes factors obtained with `bilby` and `DINGO-lensing` — $\log_{10}\mathcal{B}_{\rm lens} = 19.0$ and 18.9, respectively — are consistent within the statistical uncertainties, confirming the reliability of `DINGO-lensing` in evaluating the lensing hypothesis while reducing the inference time to ~ 40 s.

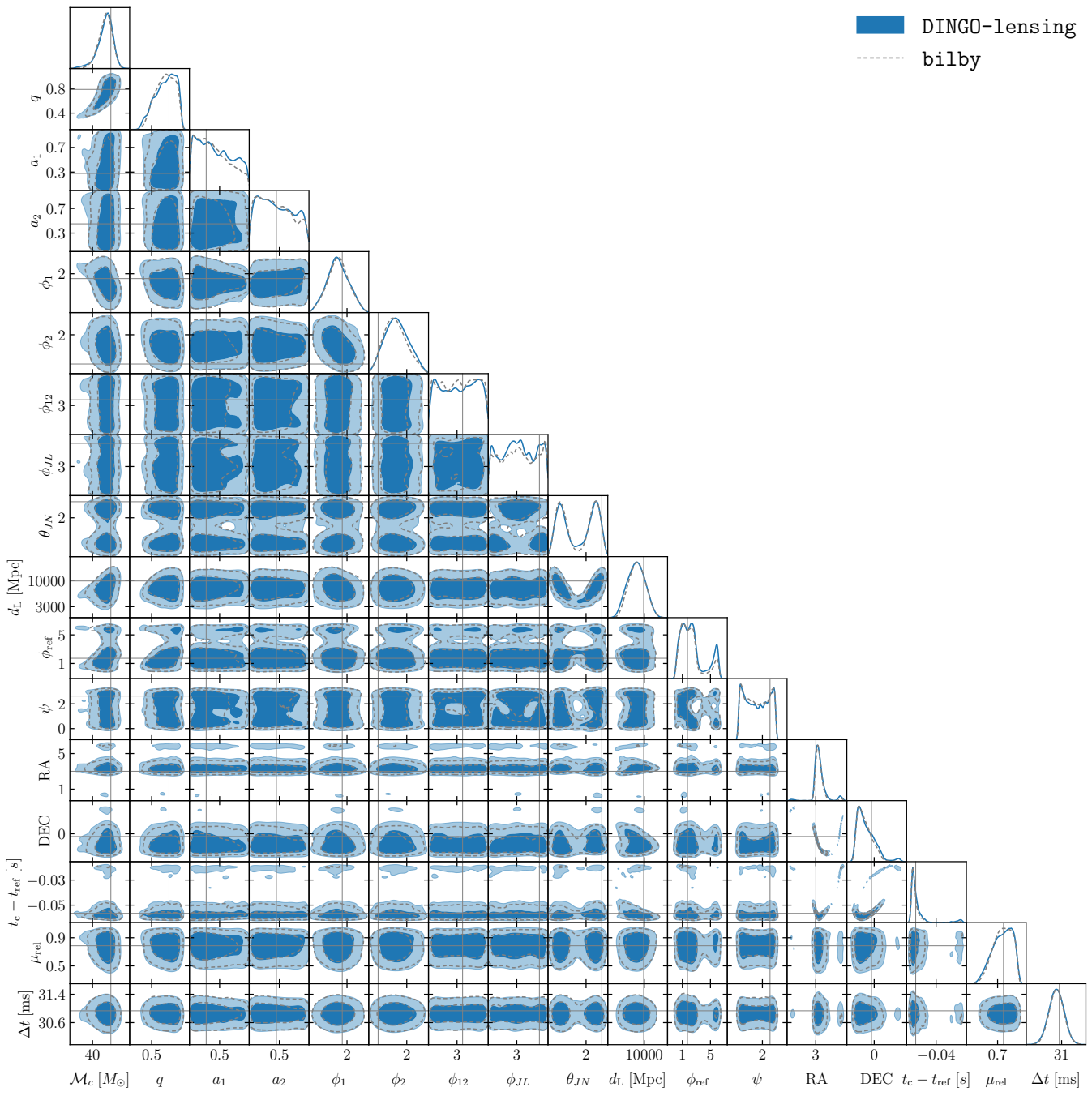

Figure 3. Posterior distribution of an injection recovered using our `DINGO-lensing` (solid blue curves) and `bilby` (dashed gray curves), respectively. The two methods agree with each other and both recovered the injected values (solid gray lines; also tabulated in Tab. I) well.

D. Recovery of GWs diffracted by point lenses

To assess the effectiveness of the geometrical optics approximation in capturing lensing distortions, we perform inference on a lensed GW signal produced by a point-mass lens for which the amplification factor is analytical and depends on the redshifted lens mass $M_{\rm Lz}$ and impact parameter y [18]. The point-mass lens always produces two images, but also includes diffraction effects which are not present in our parametrization of interfering chirps in (3). In practical terms, we use the implementation of the point-mass lens model coded in `modwaveforms`,[3]

[3] https://github.com/ezquiaga/modwaveforms

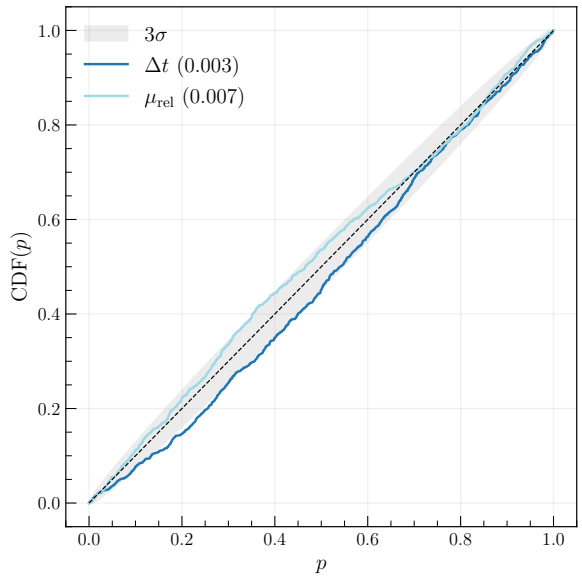

Figure 4. P-P plot showing the calibration of the posterior distributions for the lensing parameters Δt and $\mu_{\rm rel}$, obtained from 10^3 simulated injections of point-mass–lensed GW signals. Each curve compares the cumulative fraction of true parameter values enclosed within given credible levels against the ideal uniform expectation (black dashed line). The light-gray band indicates the expected 3σ statistical uncertainty for a perfectly calibrated inference. The analysis uses the geometrical optics approximation to recover signals generated with a point-lens model, demonstrating accurate recovery of both the time delay and relative magnification distributions.

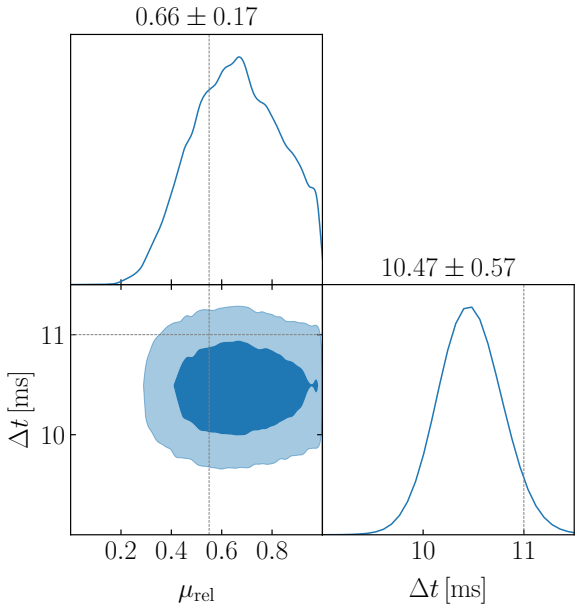

Figure 5. The injection corresponds to a point-mass lens with $M_{\rm Lz} = 900 M_\odot$ and $y = 0.3$. The inferred $\Delta t = 10.47 \pm 0.57$ ms and $\mu_{\rm rel} = 0.66 \pm 0.17$ are consistent with the geometrical optics predictions of $\Delta t = 11$ ms and $\mu_{\rm rel} = 0.5$, demonstrating the capability of the geometrical optics approximation to accurately reproduce lensing-induced distortions.

which is derived on Ref. [108]. Within the geometric optics limit, the lens mass and impact parameter can be mapped one-to-one with the lensing time delay and relative magnification:

$$\Delta t = \frac{4GM_{\rm Lz}}{c^3}\Delta T(y)\,, \tag{10}$$

$$\mu_{\rm rel} = \left|\frac{y\sqrt{y^2+4}-(y^2+2)}{y\sqrt{y^2+4}+(y^2+2)}\right|\,, \tag{11}$$

where the dimensionless time delay is determined by the impact parameter only

$$\Delta T(y) = \frac{1}{2}y\sqrt{y^2+4} - \ln\left|\frac{y-\sqrt{y^2+4}}{y-\sqrt{y^2+4}}\right|\,. \tag{12}$$

The time delay scales linearly with the lens mass and it is inversely proportional with the impact parameter. For a $100M_\odot$ lens, the time delay is on the order of $4GM_{\rm Lz}/c^3 \approx 2$ms. In the limit of $y \to 0$, then $\mu_{\rm rel} \to 1$.

Figure 4 presents the P-P plot assessing the calibration of posterior inferences for the lensing parameters Δt and $\mu_{\rm rel}$, based on 10^3 simulated injections of GW signals lensed by a point-mass lens, and 10^4 sample were drawn from the lensed network for each injection. We simulate lensed events by uniformly sampling lens masses from $500M_\odot$ to $1500M_\odot$. Impact parameters are drawn uniformly from 0 to 1.5. This choice of parameters ensures that the associated time delay does not exceed 0.1s. Each curve shows the cumulative fraction of true parameter values contained within the corresponding nominal credible intervals, compared to the ideal uniform expectation indicated by the black dashed line. The light-gray region denotes the expected 3σ statistical variation for a perfectly calibrated inference under the null hypothesis. While both Δt and $\mu_{\rm rel}$ curves exhibit mild deviations beyond the 3σ envelope, these offsets are small and do not indicate significant systematic biases. Moreover, we also include the corresponding p-values from the Kolmogorov-Smirnov test in the legend. This level of agreement demonstrates that, even when using the geometrical optics approximation to recover signals generated with a point-lens model, the inference remains well calibrated and the residual systematics are negligible relative to statistical fluctuations.

To further illustrate the quality of individual recoveries, we focus on a representative example of a lensed signal. Figure 5 presents a subset of the 17-dimensional corner plot and illustrates the recovery of a lensed GW signal using the `DINGO-lensing` framework for a point-mass lens with $M_{\rm Lz} = 900\ M_\odot$ and $y = 0.3$. 10^5 samples have been drawn from the lensed network with utilizing importance sampling. The two-dimensional posterior distributions show tight constraints on the lensing parameters, indicating that the model successfully captures the lens-

induced distortions. The inferred $\Delta t = 10.47 \pm 0.57$ ms and $\mu_{\rm rel} = 0.6 \pm 0.17$ are in excellent agreement with the geometrical optics predictions of $\Delta t = 11$ ms and $\mu_{\rm rel} = 0.5$, respectively, demonstrating the capability of the geometrical optics approximation to accurately reproduce the lensing effects encoded in the waveform.

We note that this demonstration does not rely on a fully self-consistent point-lensed source model for parameter estimation. Such models, while analytically tractable, neglect waveform modifications arising from finite-source effects and frequency-dependent interference, which become significant outside the geometrical optics regime. Instead, our approach focuses on validating the ability of the `DINGO-lensing` network to recover the effective observables (Δt, $\mu_{\rm rel}$) that characterize the lensing distortion within this limit. A direct comparison with full Bayesian recovery using `bilby` is not performed here, as the inference is expensive and other codes have already implemented this [109, 110]. Our inference test therefore serves as a controlled and efficient validation of the capability of the network to reproduce the predictions of geometrical optics with high fidelity.

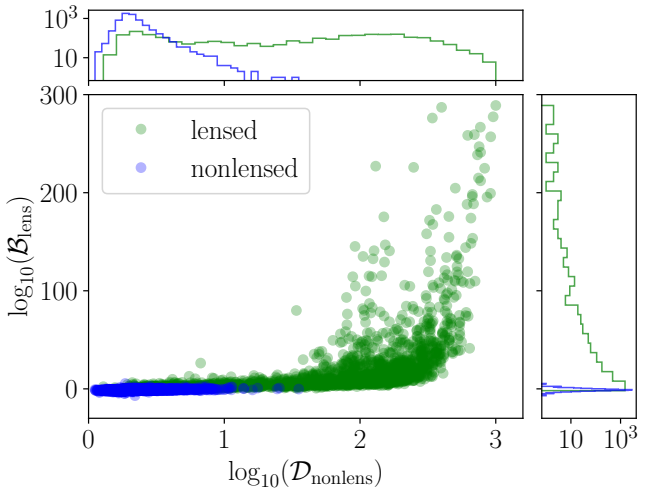

Figure 6. Distribution of lensing Bayes factors ($\mathcal{B}_{\rm lens}$) and distance away from the nonlensing hypothesis ($\mathcal{D}_{\rm nonlens}$) for the set of 2,561 lensed and 10,858 nonlensed simulated events considered in this work.

E. Lensing detectability

`DINGO-lensing` enables us to assess the detectability of gravitational waveform distortions induced by lensing. To that end, we simulate a large set of lensed and nonlensed events for our fiducial Hanford-Livingston O5 detector network. We sample events uniformly in a comoving volume between 1Gpc and 15Gpc for chirp masses between $30M_\odot$ and $60M_\odot$ and mass ratios larger than 0.3. The rest of the parameters are sampled from the fiducial distributions used in network training. The reference trigger GPS time is fixed to 100s (with a uniform scatter of $\pm$0.1s), and we only consider events with a network SNR larger than 8. For every set of sampled parameters, we generate a frame file and define the detector strain data for a given Gaussian noise realization of the power spectral density. We analyze each frame file with both our lensed and nonlensed networks in order to compute the lensing Bayes factor using importance sampling. We have validated our simulations against `bilby` for instances with different sampling efficiencies, see Appendix B, finding good agreement in the Bayes factor for $\epsilon > 10^{-4}$, which we set as a lower bound.

Lensing candidates can be identified through different methods. For example, using the posterior samples directly would give an indication of an inconsistency with the nonlensing hypothesis if the lensing parameters deviate away from 0. To quantify this inconsistency, we follow a similar approach to Ref. [111] and compute the (Gaussian) distance to $(0, 0)$ in the 2D parameter space $(\Delta t, \mu_{\rm rel})$. We define this distance away from the nonlensing hypothesis as $\mathcal{D}_{\rm nonlens}$, see appendix C for details. Such distance is weighted by the uncertainty of the lensing parameters. Well-constrained posteriors away from 0 would lead to larger distances. Because both lensing parameters are bounded at 0, distances are always larger than 1. We compare the distributions of $\mathcal{D}_{\rm nonlens}$ and $\mathcal{B}_{\rm lens}$ in Fig. 6 for 2,561 lensed and 10,858 nonlensed simulations. We find that both are correlated, with lensed simulations displaying a large tail at high values. Noticeably, the distribution of nonlensed simulations has a narrower range of values both in distance and Bayes factors.

We explore in more detail the distribution of Bayes factors in Fig. 7, where we plot the complementary cumulative distribution functions. Lensed simulations display large Bayes factor, with 53% of them being larger than 5. However, for a given lensing candidate, its statistical significance depends on how it compares to a 'background', which defines the expectations of the nonlensed hypothesis. In other words, it is necessary to compute the *false alarm probability*. Similar background studies with `DINGO` have also been performed in the literature when searching for eccentric signals [98].

We find that only 2% of nonlensed simulations favor lensing, $\mathcal{B}_{\rm lens} > 0$. The background, on the other hand, is sensitive to the underlying distributions being considered. We illustrate this in Fig. 7 by distinguishing between nonlensed events with chirp masses in the ranges $[30,40]M_\odot$ and $[40,60]M_\odot$. We find that nonlensed events with larger masses tend to mimic lensing at a higher rate. If we consider only the subset of nonlensed simulations with either of the dimensionless, component spin magnitudes ($a_{1,2}$) larger than 0.5, we find that the false alarm rate also increases. This matches with the expectation that highly spinning signals could produce larger precession and be shorter in duration, being more likely to resemble lensing by random chance.

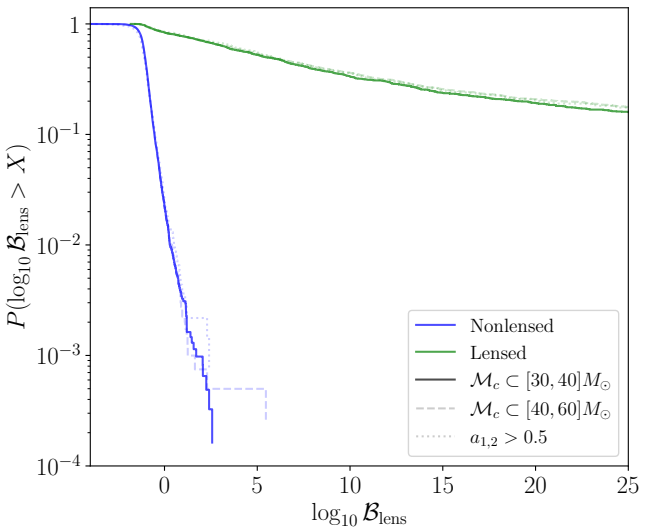

Figure 7. Complementary cumulative distribution function of lensing Bayes factors ($\mathcal{B}_{\rm lens}$) for simulated lensed and nonlensed with O5 sensitivity. We distinguish between events with chirp masses between [30,40]$M_\odot$ (solid lines) and [40,60]$M_\odot$ (dashed lines). For all chirp masses, we also plot the cases with both component dimensionless spin magnitudes smaller than 0.5 (dotted lines).

IV. CONCLUSIONS

In this work, we introduce `DINGO-lensing`, a simulation-based inference framework for performing fast and accurate inference on gravitationally lensed GW signals with deep learning approaches. `DINGO-lensing` bypasses the computational bottlenecks of conventional samplers for Bayesian inference of distorted GW from lensing while maintaining statistical interpretability.

Using a dataset of 5×10^6 simulated waveforms, we have trained lensed and nonlensed models capable of producing well-calibrated posteriors, as validated through P-P plots and comparisons with standard parameter estimation analyses with `bilby` [112]. The results demonstrate excellent agreement across intrinsic, extrinsic, and lensing parameters, while reducing the inference time from weeks to tens of seconds per event.

We demonstrate the potential of `DINGO-lensing` by training networks with LIGO detectors at design sensitivity. We find that the lensing parameters can be accurately recovered, with precision in the lensing time delay of the order of milliseconds. We show that our fiducial lens model with two overlapping chirps is able to recover lensed GWs diffracted by point-mass lenses. Moreover, we assess the detectability of lensed GWs by simulating thousands of lensed and nonlensed events. We find that for typical GW signals, the distribution of lensing Bayes factors of lenses and nonlensed events is clearly distinguishable. However, we also find that the statistical significance of a given lensed event is sensitive to the background it is compared to. In particular, we observe a dependence in mass and spin. When comparing a signal with a background of heavier and more spinning binaries, the required Bayes factors to arrive at the same false alarm probability is larger than when comparing with lighter and less spinning signals. A detailed detectability forecast will be addressed in future work.

The discovery of lensed GWs requires well-defined detection strategies. `DINGO-lensing` can contribute to them in two different stages. Firstly, it can be used for the identification of lensed candidates with a set of networks trained on the average power spectral density over a given observing run. These networks should encompass broad mass ranges characterizing different signals duration. The maximum lensing time delay should be adjusted for each network so that there are single, distorted lensed waveforms. Lensing candidates can then be identified through two alternative methods. On one hand, one can compute the distance between the lensing posterior samples and the nonlensing hypothesis, $\Delta t = \mu_{\rm rel} = 0$ using only neural posterior estimation. On the other hand, through the use of importance sampling, the Bayes factor between the lensed and nonlensed hypothesis can be compared to pre-computed simulations of nonlensed events. To avoid potential biases, such simulations should describe agnostic models, i.e. uniformly distributed priors in the networks, and astrophysically motivated ones, i.e. matching GW population models with lensing optical depths.

Secondly, `DINGO-lensing` can be used to train specialized networks for the most significant candidates. This second step is essential to account for the non-stationarity of GW detector sensitivities and the best-fit waveform approximant and signal properties. The specialized network then provides the means to compute a dedicated background to assess the statistical significance of a particular lensed candidate, which would require thousands/millions of simulations to arrive at $3\sigma/5\sigma$ detections. Obtaining a high statistical significance is a necessary step before drawing any astrophysical implication.

Altogether, `DINGO-lensing` provides a scalable pathway for rapid lensing analyses in upcoming LVK observing runs. Its efficiency enables large-scale background estimation and systematic population studies that are otherwise computationally prohibitive. Future work will extend this framework to more complex lens models, incorporate waveform systematics, and integrate `DINGO-lensing` into end-to-end pipelines for real-time identification and characterization of lensed GW events.

ACKNOWLEDGMENTS

We thank Srashti Goyal, Miguel Zumalacárregui, Nihar Gupte, Annalena Kofler, Stephen R. Green, Alvin Li, Elwin Ka Yau Li, and Arthur Offermans for valuable discussions and insightful comments. We also thank David Keitel for carefully reviewing this article for

LVK, providing many constructive comments. This work was supported by Research Grants No. VIL37766 and No. VIL53101 from Villum Fonden and the DNRF Chair Program Grant No. DNRF162 by the Danish National Research Foundation. This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 101131233. J. M. E is also supported by the Marie Skłodowska-Curie Grant Agreement No. 847523 INTERACTIONS. The Center of Gravity is a Center of Excellence funded by the Danish National Research Foundation under grant no. DNRF184. The Tycho supercomputer hosted at the SCIENCE HPC center at the University of Copenhagen was used for supporting this work.

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.

DATA AVAILABILITY

`DINGO-lensing` is available on GitHub [113]. The neural networks used in this work will be made public upon publication.

Note added.—During the preparation of the manuscript, we learnt that Caldarola et al. have similarly explored using `DINGO` for analyzing signals lensed by a point mass [114].

[1] J. Aasi *et al.* (LIGO Scientific), Class. Quant. Grav. **32**, 074001 (2015), arXiv:1411.4547 [gr-qc].
[2] F. Acernese *et al.* (VIRGO), Class. Quant. Grav. **32**, 024001 (2015), arXiv:1408.3978 [gr-qc].
[3] T. Akutsu *et al.* (KAGRA), PTEP **2021**, 05A101 (2021), arXiv:2005.05574 [physics.ins-det].
[4] A. G. Abac et al. (LIGO Scientific, VIRGO, KAGRA), (2025), arXiv:2508.18080 [gr-qc].
[5] S. E. Field, C. R. Galley, J. S. Hesthaven, J. Kaye, and M. Tiglio, Phys. Rev. X **4**, 031006 (2014), arXiv:1308.3565 [gr-qc].
[6] A. Nagar *et al.*, Phys. Rev. D **98**, 104052 (2018), arXiv:1806.01772 [gr-qc].
[7] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, L. E. Kidder, and H. P. Pfeiffer, Phys. Rev. D **99**, 064045 (2019), arXiv:1812.07865 [gr-qc].
[8] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, Phys. Rev. Research. **1**, 033015 (2019), arXiv:1905.09300 [gr-qc].
[9] G. Pratten *et al.*, Phys. Rev. D **103**, 104056 (2021), arXiv:2004.06503 [gr-qc].
[10] S. Ossokine *et al.*, Phys. Rev. D **102**, 044055 (2020), arXiv:2004.09442 [gr-qc].
[11] E. Hamilton, L. London, J. E. Thompson, E. Fauchon-Jones, M. Hannam, C. Kalaghatgi, S. Khan, F. Pannarale, and A. Vano-Vinuales, Phys. Rev. D **104**, 124027 (2021), arXiv:2107.08876 [gr-qc].
[12] L. Pompili *et al.*, Phys. Rev. D **108**, 124035 (2023), arXiv:2303.18039 [gr-qc].
[13] A. Ramos-Buades, A. Buonanno, H. Estellés, M. Khalil, D. P. Mihaylov, S. Ossokine, L. Pompili, and M. Shiferaw, Phys. Rev. D **108**, 124037 (2023), arXiv:2303.18046 [gr-qc].
[14] J. E. Thompson, E. Hamilton, L. London, S. Ghosh, P. Kolitsidou, C. Hoy, and M. Hannam, Phys. Rev. D **109**, 063012 (2024), arXiv:2312.10025 [gr-qc].
[15] M. Colleoni, F. A. R. Vidal, C. García-Quirós, S. Akçay, and S. Bera, Phys. Rev. D **111**, 104019 (2025), arXiv:2412.16721 [gr-qc].
[16] S. Albanesi, R. Gamba, S. Bernuzzi, J. Fontbuté, A. Gonzalez, and A. Nagar, (2025), arXiv:2503.14580 [gr-qc].
[17] T. T. Nakamura, Phys. Rev. Lett. **80**, 1138 (1998).
[18] R. Takahashi and T. Nakamura, Astrophys. J. **595**, 1039 (2003), arXiv:astro-ph/0305055.
[19] P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational Lenses*, Astronomy and Astrophysics Library (Springer, 1992).
[20] R. K. L. Lo, L. Vujeva, J. M. Ezquiaga, and J. C. L. Chan, Phys. Rev. Lett. **134**, 151401 (2025), arXiv:2407.17547 [gr-qc].
[21] L. Vujeva, J. M. Ezquiaga, R. K. L. Lo, and J. C. L. Chan, Phys. Rev. D **112**, 063044 (2025), arXiv:2501.02096 [astro-ph.CO].
[22] J. M. Ezquiaga, R. K. L. Lo, and L. Vujeva, Phys. Rev. D **112**, 043544 (2025), arXiv:2503.22648 [gr-qc].
[23] L. Vujeva, J. M. Ezquiaga, D. Gilman, S. Goyal, and M. Zumalacárregui, (2025), arXiv:2510.14953 [astro-ph.CO].
[24] A. Liu and K. Kim, Phys. Rev. D **110**, 123008 (2024), arXiv:2301.07253 [gr-qc].

[25] N. Rao, A. Mishra, A. Ganguly, and A. More, (2025), arXiv:2510.17787 [gr-qc].
[26] J. C. L. Chan, E. Seo, A. K. Y. Li, H. Fong, and J. M. Ezquiaga, Phys. Rev. D **111**, 084019 (2025), arXiv:2411.13058 [gr-qc].
[27] S. Jung and C. S. Shin, Phys. Rev. Lett. **122**, 041103 (2019), arXiv:1712.01396 [astro-ph.CO].
[28] K.-H. Lai, O. A. Hannuksela, A. Herrera-Martín, J. M. Diego, T. Broadhurst, and T. G. Li, Phys. Rev. D **98**, 083005 (2018), arXiv:1801.07840 [gr-qc].
[29] P. Christian, S. Vitale, and A. Loeb, Phys. Rev. D **98**, 103022 (2018), arXiv:1802.02586 [astro-ph.HE].
[30] L. Dai, S.-S. Li, B. Zackay, S. Mao, and Y. Lu, Phys. Rev. D **98**, 104029 (2018), arXiv:1810.00003 [gr-qc].
[31] J. Diego, O. Hannuksela, P. Kelly, T. Broadhurst, K. Kim, T. Li, G. Smoot, and G. Pagano, Astron. Astrophys. **627**, A130 (2019), arXiv:1903.04513 [astro-ph.CO].
[32] J. M. Diego, Phys. Rev. D **101**, 123512 (2020), arXiv:1911.05736 [astro-ph.CO].
[33] K. Liao, S. Tian, and X. Ding, Mon. Not. Roy. Astron. Soc. **495**, 2002 (2020), arXiv:2001.07891 [astro-ph.CO].
[34] A. Mishra, A. K. Meena, A. More, S. Bose, and J. S. Bagla, Mon. Not. Roy. Astron. Soc. **508**, 4869 (2021), arXiv:2102.03946 [astro-ph.CO].
[35] E. Seo, O. A. Hannuksela, and T. G. F. Li, Astrophys. J. **932**, 50 (2022), arXiv:2110.03308 [astro-ph.HE].
[36] E. Seo, T. G. F. Li, and M. A. Hendry, Astrophys. J. **966**, 107 (2024), arXiv:2311.05543 [gr-qc].
[37] A. Liu, I. C. F. Wong, S. H. W. Leong, A. More, O. A. Hannuksela, and T. G. F. Li, Mon. Not. Roy. Astron. Soc. **525**, 4149 (2023), arXiv:2302.09870 [gr-qc].
[38] M. H.-Y. Cheung, K. K. Y. Ng, M. Zumalacárregui, and E. Berti, Phys. Rev. D **109**, 124020 (2024), arXiv:2403.13876 [gr-qc].
[39] M. Zumalacárregui, (2024), arXiv:2404.17405 [gr-qc].
[40] A. Chakraborty and S. Mukherjee, (2024), arXiv:2410.06995 [gr-qc].
[41] J. C. L. Chan, C. Dyson, M. Garcia, J. Redondo-Yuste, and L. Vujeva, Phys. Rev. D **112**, 064009 (2025), arXiv:2502.14073 [gr-qc].
[42] M. Fairbairn, J. Urrutia, and V. Vaskonen, JCAP **07**, 007 (2023), arXiv:2210.13436 [astro-ph.CO].
[43] G. Tambalo, M. Zumalacárregui, L. Dai, and M. H.-Y. Cheung, Phys. Rev. D **108**, 103529 (2023), arXiv:2212.11960 [astro-ph.CO].
[44] M. Çalışkan, L. Ji, R. Cotesta, E. Berti, M. Kamionkowski, and S. Marsat, Phys. Rev. D **107**, 043029 (2023), arXiv:2206.02803 [astro-ph.CO].
[45] S. Savastano, G. Tambalo, H. Villarrubia-Rojo, and M. Zumalacarregui, Phys. Rev. D **108**, 103532 (2023), arXiv:2306.05282 [gr-qc].
[46] S. Kim, H. Gil Choi, and S. Jung, JHEP **07**, 006 (2025), arXiv:2501.14904 [hep-ph].
[47] K. T. Abe, S. Arai, R. Inui, T. S. Yamamoto, H. Yarimoto, and S. Yokoyama, (2025), arXiv:2506.07507 [astro-ph.CO].
[48] S. Ali, M. Kesden, and L. King, (2025), arXiv:2506.18750 [gr-qc].
[49] Z. Su, X. Shan, Z. Lyu, J. Zhang, Y. Liu, S. Mao, and H. Yang, (2025), arXiv:2510.17125 [gr-qc].
[50] O. Bulashenko, N. Villanueva, R. B. Nerin, and J. A. Font, (2025), arXiv:2510.20442 [gr-qc].
[51] S. Singh, G. Brando de Oliveira, S. Savastano, and M. Zumalacárregui, JCAP **07**, 025 (2025), arXiv:2502.10758 [astro-ph.CO].
[52] P. Cremonese, J. M. Ezquiaga, and V. Salzano, Phys. Rev. D **104**, 023503 (2021), arXiv:2104.07055 [astro-ph.CO].
[53] A. Chen, P. Cremonese, J. M. Ezquiaga, and D. Keitel, (2024), arXiv:2408.03856 [astro-ph.CO].
[54] G. P. Smith *et al.*, Phil. Trans. Roy. Soc. Lond. A **383**, 20240134 (2025), arXiv:2503.19973 [astro-ph.HE].
[55] S. H. W. Leong, J. Janquart, A. K. Sharma, P. Martens, P. Ajith, and O. A. Hannuksela, Astrophys. J. Lett. **979**, L27 (2025), arXiv:2408.13144 [astro-ph.HE].
[56] Y.-Z. Li, W.-L. Xu, Y.-G. Chen, and W.-H. Lei, (2025), arXiv:2506.04196 [astro-ph.HE].
[57] H. Ubach, M. Gieles, and J. Miralda-Escudé, Phys. Rev. D **112**, 083026 (2025), arXiv:2505.04794 [astro-ph.HE].
[58] S. Goyal, K. Haris, A. K. Mehta, and P. Ajith, Phys. Rev. D **103**, 024038 (2021), arXiv:2008.07060 [gr-qc].
[59] S. Goyal, A. Vijaykumar, J. M. Ezquiaga, and M. Zumalacarregui, Phys. Rev. D **108**, 024052 (2023), arXiv:2301.04826 [gr-qc].
[60] H. Narola, J. Janquart, L. Haegel, K. Haris, O. A. Hannuksela, and C. Van Den Broeck, Phys. Rev. D **109**, 084064 (2024), arXiv:2308.01709 [gr-qc].
[61] J. M. Ezquiaga and M. Zumalacárregui, Phys. Rev. D **102**, 124048 (2020), arXiv:2009.12187 [gr-qc].
[62] A. Liu, R. S. Chandramouli, O. A. Hannuksela, N. Yunes, and T. G. F. Li, (2024), arXiv:2410.21738 [gr-qc].
[63] D. Keitel, Phil. Trans. Roy. Soc. Lond. A **383**, 20240128 (2025), arXiv:2407.12974 [gr-qc].
[64] O. A. Hannuksela, K. Haris, K. K. Y. Ng, S. Kumar, A. K. Mehta, D. Keitel, T. G. F. Li, and P. Ajith, Astrophys. J. Lett. **874**, L2 (2019), arXiv:1901.02674 [gr-qc].
[65] L. Dai, B. Zackay, T. Venumadhav, J. Roulet, and M. Zaldarriaga, (2020), arXiv:2007.12709 [astro-ph.HE].
[66] R. Abbott *et al.* (LIGO Scientific, VIRGO), Astrophys. J. **923**, 14 (2021), arXiv:2105.06384 [gr-qc].
[67] R. Abbott *et al.* (LIGO Scientific, VIRGO, KAGRA), arXiv e-prints , arXiv:2304.08393 (2023), arXiv:2304.08393 [gr-qc].
[68] J. Janquart *et al.*, Mon. Not. Roy. Astron. Soc. **526**, 3832 (2023), arXiv:2306.03827 [gr-qc].
[69] A. Chakraborty and S. Mukherjee, Astrophys. J. **990**, 68 (2025), arXiv:2503.16281 [gr-qc].
[70] A. G. Abac *et al.* (LIGO Scientific, VIRGO, KAGRA), (2025), arXiv:2507.08219 [astro-ph.HE].
[71] Juno C. L. Chan, J. M. Ezquiaga, R. K. L. Lo, J. Bowman, L. Magaña Zertuche, and L. Vujeva, (to appear) (2025).
[72] E. Thrane and C. Talbot, Publ. Astron. Soc. Austral. **36**, e010 (2019), [Erratum: Publ.Astron.Soc.Austral. 37, e036 (2020)], arXiv:1809.02293 [astro-ph.IM].
[73] B. Liang and H. Wang, (2025), 10.61977/ati2025020, arXiv:2507.11192 [gr-qc].
[74] E. Seo, X. Shan, J. Janquart, O. A. Hannuksela, M. A. Hendry, and B. Hu, Astrophys. J. **988**, 159 (2025), arXiv:2503.02186 [gr-qc].
[75] Z. Qin, T.-Y. Sun, B.-Y. Li, J.-F. Zhang, X. Guo, and X. Zhang, (2025), arXiv:2505.20996 [gr-qc].
[76] A. J. K. Chua and M. Vallisneri, Phys. Rev. Lett. **124**,

041102 (2020), arXiv:1909.05966 [gr-qc].
[77] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith, Nature Phys. **18**, 112 (2022), arXiv:1909.06296 [astro-ph.IM].
[78] E. Cuoco *et al.*, Mach. Learn. Sci. Tech. **2**, 011002 (2021), arXiv:2005.03745 [astro-ph.HE].
[79] J. Langendorff, A. Kolmus, J. Janquart, and C. Van Den Broeck, Phys. Rev. Lett. **130**, 171402 (2023), arXiv:2211.15097 [gr-qc].
[80] T. Zhao, R. Shi, Y. Zhou, Z. Cao, and Z. Ren, Front. Phys. (Beijing) **20**, 45301 (2025), arXiv:2311.15585 [gr-qc].
[81] U. Bhardwaj, J. Alvey, B. K. Miller, S. Nissanke, and C. Weniger, Phys. Rev. D **108**, 042004 (2023), arXiv:2304.02035 [gr-qc].
[82] E. Cuoco, M. Cavaglià, I. S. Heng, D. Keitel, and C. Messenger, Living Rev. Rel. **28**, 2 (2025), arXiv:2412.15046 [gr-qc].
[83] Q. Hu, J. Irwin, Q. Sun, C. Messenger, L. Suleiman, I. S. Heng, and J. Veitch, (2024), arXiv:2412.03454 [gr-qc].
[84] A. Polanska, T. Wouters, P. T. H. Pang, K. K. W. Wong, and J. D. McEwen, in *38th conference on Neural Information Processing Systems* (2024) arXiv:2410.21076 [astro-ph.IM].
[85] N. Stergioulas (2024) arXiv:2401.07406 [gr-qc].
[86] L. Papalini, F. De Santi, M. Razzano, I. S. Heng, and E. Cuoco, Class. Quant. Grav. **42**, 185012 (2025), arXiv:2505.02773 [gr-qc].
[87] F. Santoliquido *et al.*, (2025), arXiv:2504.21087 [astro-ph.HE].
[88] K. Kim, J. Lee, R. S. H. Yuen, O. A. Hannuksela, and T. G. F. Li, Astrophys. J. **915**, 119 (2021), arXiv:2010.12093 [gr-qc].
[89] S. Goyal, H. D., S. J. Kapadia, and P. Ajith, Phys. Rev. D **104**, 124057 (2021), arXiv:2106.12466 [gr-qc].
[90] S. Magare, A. More, and S. Choudhary, Mon. Not. Roy. Astron. Soc. **535**, 990 (2024), arXiv:2403.02994 [astro-ph.HE].
[91] R. Bada-Nerin, O. Bulashenko, O. Gramaxo Freitas, and J. A. Font, Phys. Rev. D **111**, 084067 (2025), arXiv:2412.00566 [gr-qc].
[92] S. R. Green, C. Simpson, and J. Gair, Phys. Rev. D **102**, 104057 (2020), arXiv:2002.07656 [astro-ph.IM].
[93] S. R. Green and J. Gair, Mach. Learn. Sci. Tech. **2**, 03LT01 (2021), arXiv:2008.03312 [astro-ph.IM].
[94] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, Phys. Rev. Lett. **127**, 241103 (2021), arXiv:2106.12594 [gr-qc].
[95] M. Dax, S. R. Green, J. Gair, M. Deistler, B. Schölkopf, and J. H. Macke, (2021), arXiv:2111.13139 [cs.LG].
[96] J. Wildberger, M. Dax, S. R. Green, J. Gair, M. Pürrer, J. H. Macke, A. Buonanno, and B. Schölkopf, Phys. Rev. D **107**, 084046 (2023), arXiv:2211.08801 [gr-qc].
[97] M. Dax, S. R. Green, J. Gair, M. Pürrer, J. Wildberger, J. H. Macke, A. Buonanno, and B. Schölkopf, Phys. Rev. Lett. **130**, 171403 (2023), arXiv:2210.05686 [gr-qc].
[98] N. Gupte *et al.*, (2024), arXiv:2404.14286 [gr-qc].
[99] Q. Hu, (2025), arXiv:2507.05209 [gr-qc].
[100] K. Cranmer, J. Brehmer, and G. Louppe, Proc. Nat. Acad. Sci. **117**, 30055 (2020), arXiv:1911.01429 [stat.ML].
[101] B. P. Abbott *et al.* (LIGO Scientific, Virgo, KAGRA), Living Rev. Rel. **21**, 3 (2018), arXiv:1304.0670 [gr-qc].
[102] https://dcc.ligo.org/LIGO-T2000012/public.
[103] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration, "LVK Algorithm Library - LALSuite," Free software (GPL) (2018).
[104] K. Wette, SoftwareX **12**, 100634 (2020).
[105] A. M. Serra and O. Bulashenko, Phys. Rev. D **112**, 024041 (2025), arXiv:2504.10128 [gr-qc].
[106] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," (2017), arXiv:1412.6980 [cs.LG].
[107] I. Loshchilov and F. Hutter, arXiv e-prints , arXiv:1608.03983 (2016), arXiv:1608.03983 [cs.LG].
[108] J. M. Ezquiaga, W. Hu, and M. Lagos, Phys. Rev. D **102**, 023531 (2020), arXiv:2005.10702 [astro-ph.CO].
[109] M. Wright and M. Hendry, The Astrophysical Journal **935**, 68 (2022), arXiv:2112.07012 [astro-ph.HE].
[110] (2023), `GWMAT`: Gravitational Waves Microlensing Analysis Tools. https://git.ligo.org/anuj.mishra/gwmat.
[111] J. M. Ezquiaga, W. Hu, and R. K. L. Lo, Phys. Rev. D **108**, 103520 (2023), arXiv:2308.06616 [astro-ph.CO].
[112] G. Ashton *et al.*, Astrophys. J. Suppl. Ser. **241**, 27 (2019), arXiv:1811.02042 [astro-ph.IM].
[113] (2025), `DINGO-lensing`: Fast parameter estimations of lensed gravitational wave signals with DINGO. https://github.com/dingo-lensing/dingo-lensing.
[114] M. Caldarola, S. Goyal, N. Gupte, S. Green, and M. Zumalacarregui, to appear (2025).

Appendix A: Nonlensed Network Validation

We train the nonlensed network following the setup and prior distribution described in the main text and Table I, except that the lensing parameters are excluded from the prior. Our architecture comprises $\sim 10^8$ learnable parameters and $\sim 10^7$ fixed parameters.

Figure 8 presents the evolution of the loss function over the course of training. Throughout optimization, the training (dashed) and validation (solid) losses remain closely aligned, indicating that the model generalizes well without signs of overfitting. Both curves show a smooth, monotonic decline, reflecting a stable and well-behaved training process.

Around epoch 300, a distinct kink appears, corresponding to the transition to Stage 1, when the previously frozen layers are unfrozen for fine-tuning. The loss subsequently stabilizes after roughly 400 epochs, implying convergence and the saturation of performance gains. Training is therefore halted at epoch 450, once no appreciable improvement in the validation loss is observed. The entire training procedure requires approximately 10 days on an NVIDIA A100 PCIe GPU.

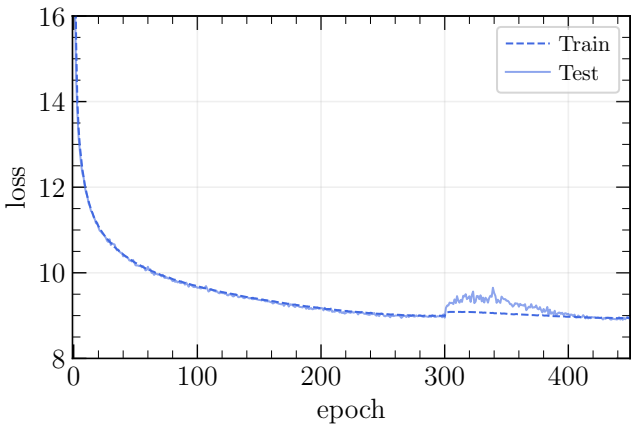

Figure 8. The loss as a function of the number of training epochs of the nonlensed network, shown for both the training set (dashed) and testing set (solid).

Figure 9 displays the P-P plot obtained from the posterior distributions of the nonlensed injection set. The black diagonal indicates the ideal expectation that $x\%$ of the true parameter values fall within the corresponding $x\%$ credible intervals. The colored curves representing individual parameters follow this diagonal closely, signifying that the inferred posteriors are statistically well calibrated. Such consistency demonstrates that the inference network is accurately able to cover the true parameter space. Furthermore, as seen by the p-values reported in the legend, the Kolmogorov–Smirnov test confirms the reliability of the sampling procedure.

Figure 10 shows a corner plot comparing the posterior distributions from the `DINGO` nonlensed model (solid

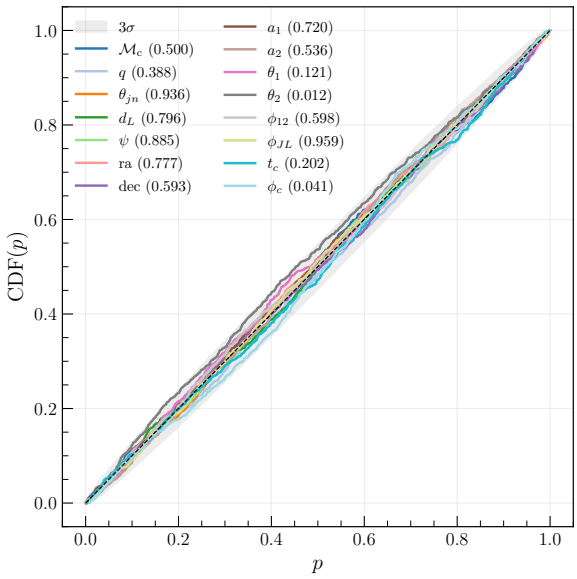

Figure 9. P-P plot for the nonlensed neural posterior estimation model constructed using 10^3 unlensed injections and no importance sampling. For each injection, a posterior is generated, and the percentile value of the injected parameter is computed. Each colored line corresponds to the cumulative distribution function (CDF) of an individual parameter, with the Kolmogorov–Smirnov (KS) test p-values given in the legend.

green contours) with those from the reference `bilby` analysis (dashed gray contours) for the same *lensed* injection used in Sec. III C. The true parameters and the priors adopted in the `bilby` run are listed in Table I. The one- and two-dimensional marginalized posteriors exhibit excellent consistency between `DINGO` and `bilby` across both intrinsic and extrinsic parameters, confirming that the neural posterior estimator faithfully reproduces the full Bayesian inference. The sampling efficiency for `DINGO` here is 0.3%.

Appendix B: Bayes factor dependence on the sampling efficiency

To further evaluate the efficiency of the posterior reconstruction, we examine the sampling efficiency ϵ obtained after applying importance sampling in our simulations validating the lensed and nonlensed networks. We perform inference on 10^3 lensed injections and 10^5 samples was drawn for each injection from the lensed network. The results for the lensed network are shown in Fig. 11. The top panel shows the histogram of $\log_{10}(\epsilon)$, with samples distributed between 10^{-1} and 10^{-4}. The middle panel displays the distribution in the $(\mathcal{M}_c, d_L)$ plane, which follows a trend of higher ϵ at larger d_L, indicating that the lensed network learns more effectively at lower SNR. This is expected as posteriors become broader and it is easier for distributions to overlap. The bottom

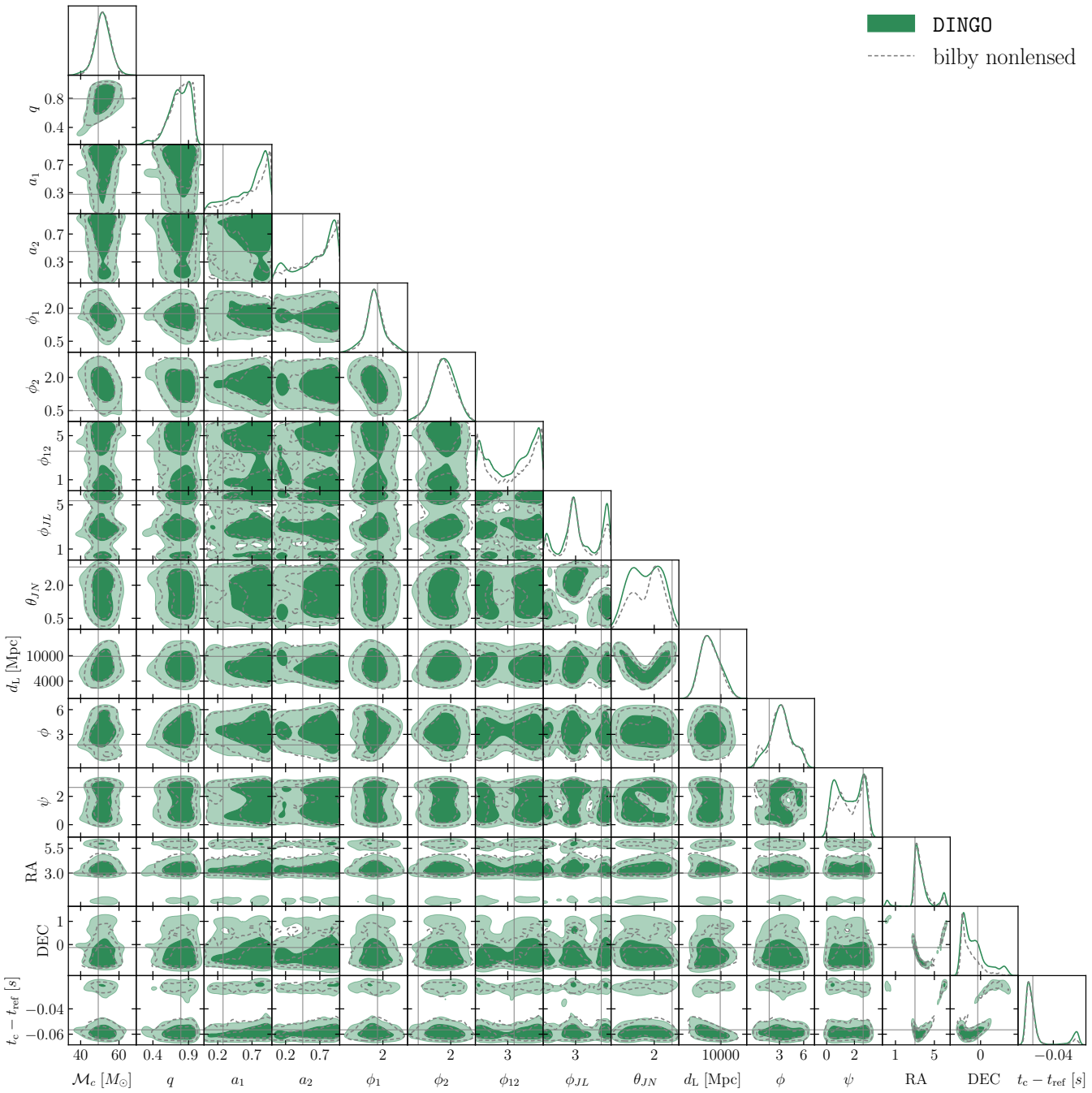

Figure 10. Posterior distribution of an injection recovered using `DINGO` (nonlensed network) (green solid curves) and `bilby` (gray dashed curves), respectively. The two methods agree with each other and both recovered the injected values (solid gray lines; also tabulated in Tab. I) well.

panel demonstrates the distribution of $\log_{10}(\epsilon)$ in the lens parameter space and reveals that the $\log_{10}(\epsilon)$ shows no significant dependence on the lensing parameters.

Figure 12 presents the distribution of the ϵ obtained from the nonlensed inference network. We perform inference 10^3 nonlensed injections and 10^5 samples was drawn for each injection from the nonlesed network. The upper panel shows the histogram of $\log_{10}(\epsilon)$, indicating that most posterior samples achieve efficiencies between $10^{-2.5}$ and 10^{-1}, with a tail extending toward lower values. The lower panel illustrates how the efficiency varies across the source parameter space, where each point corresponds to an injection colored by $\log_{10}(\epsilon)$ in the $(\mathcal{M}_c, d_L)$ plane. A mild trend is observed in which higher sampling efficiency is associated with systems of larger chirp mass and greater luminosity distance, sug-

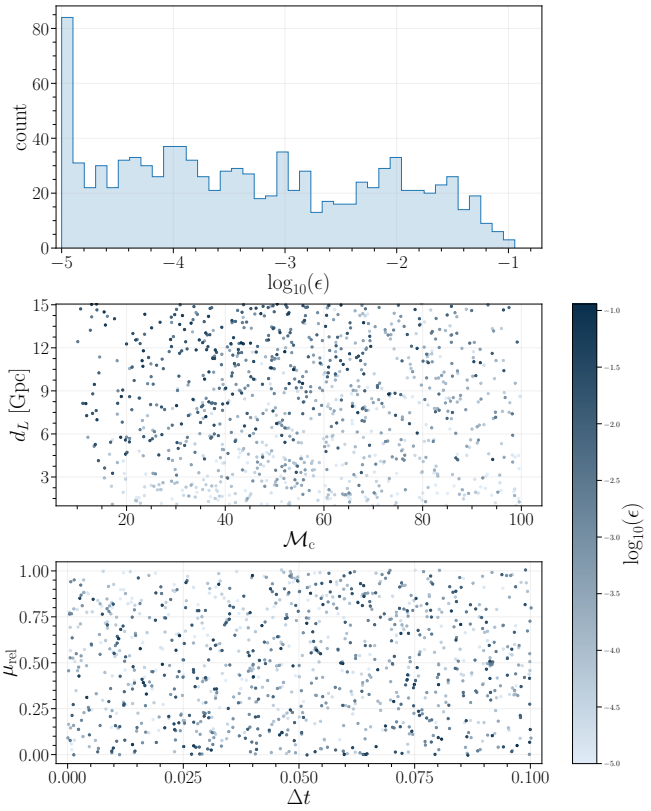

Figure 11. Distribution of the sampling efficiencies ϵ of 10^3 lensed injections for the lensed network. 10^5 samples are drawn for each injections The top panel shows the histogram of $\log_{10}(\epsilon)$, with most samples distributed between 10^{-4} and 10^{-1}. The middle panel illustrates the dependence of ϵ on the source parameters, where each point represents an injection colored by $\log_{10}(\epsilon)$ in the $(\mathcal{M}_c, d_L)$ plane. A weak trend is observed in which higher $\mathcal{M}_c$ and larger d_L correspond to an improved sampling efficiency. The bottom panel shows the variation of ϵ with the lensing parameters $(\Delta t, \mu_{rel})$, indicating no strong correlation but a broad spread consistent with the added complexity of the lensing parameter space.

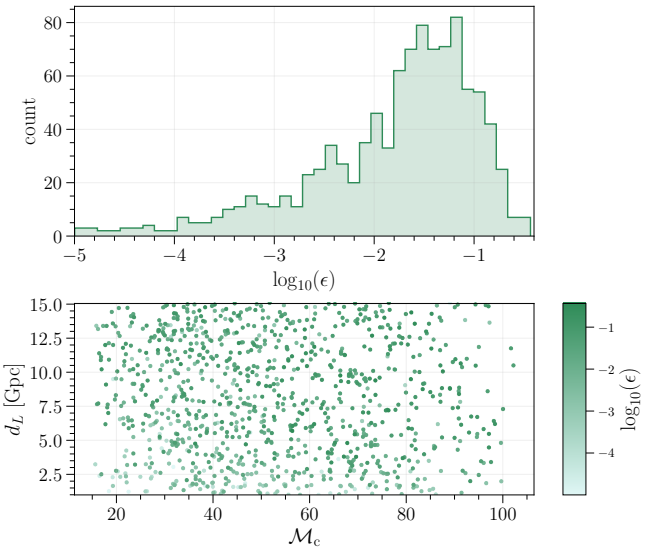

Figure 12. Distribution of the sampling efficiency ϵ of 10^3 nonlensed injections for the nonlensed network. 10^5 samples are drawn from the nonlensed network. The top panel shows the histogram of $\log_{10}(\epsilon)$, illustrating that most posterior samples have efficiencies between $10^{-2.5}$ and 10^{-1}. The bottom panel presents the relation between luminosity distance d_L and chirp mass $\mathcal{M}_c$, color-coded by $\log_{10}(\epsilon)$. The results indicate that lower-efficiency samples are more prevalent at small d_L and moderate $\mathcal{M}_c$, while the overall spread remains broad across the parameter space.

gesting that the nonlensed network more effectively reconstructs at lower-SNR and massive binaries. Overall, the nonlensed network exhibits stable performance across the population, with variations in ϵ primarily reflecting differences in signal strength rather than systematic inference bias.

Our main metric to evaluate lensed candidates is the lensing Bayes factor $\mathcal{B}_{\rm lens}$. To validate the performance of our network, we randomly select lensed and nonlensed events from our sets of simulations with different `DINGO-lensing` Bayes factors and sampling efficiencies and compare them with `bilby`. We focus on cases with $\mathcal{B}_{\rm lens} > 0$. Simulated events with `DINGO-lensing` have 5×10^4 samples. The results are summarized in table II. We observe that the Bayes factors are consistent with each other within statistical uncertainties for $\epsilon > 10^{-4}$. Therefore, we cut our simulations below that value, to end up with 2,561 and 10,858 lensed and nonlensed simulations, respectively.

During this validation, we have observed that the networks sometimes struggle to compute the evidences when the simulated mass ratios are close to the lower bound of the training set, $q = 0.125$. This may be due to a difficulty in learning the sharp features of a railing posterior distribution. For this reason, we conservatively limit our background/foreground simulations to $q > 0.3$ where we see good agreement as demonstrated above. The specific parameters of our validation set are presented in Table III.

Appendix C: Distance away from the nonlensing hypothesis

For our lens model, cf. Eq. (3), a nonlensed event is characterized by having $\Delta t = \mu_{\rm rel} = 0$. Therefore, one can quantify the inconsistency with this hypothesis by measuring the distance of the lensing posterior samples away from this point. We define this quantity as $\mathcal{D}_{\rm nonlens}$. In order for this distance to be meaningful, it needs to be normalized by the uncertainty of the parameters themselves. Approximating the posterior distributions as Gaussian, the distance in units of σ is given by [111]

$$D(\theta_1, \theta_2) = \sqrt{(\theta_1 - \theta_2))^T (C_1 + C_2)^{-1} (\theta_1 - \theta_2)}\,, \quad \text{(C1)}$$

ID	$\log_{10}\mathcal{B}^{\texttt{DINGO}}_{\text{lens}}$	$\log_{10}\mathcal{B}^{\texttt{bilby}}_{\text{lens}}$	$\epsilon_{\text{lens}} \times 10^3$	$\epsilon_{\text{nonlens}} \times 10^3$
6	1.24	-0.13	0.02	0.62
594	0.07	0.10	0.12	8.43
779	1.20	1.83	25.29	106.16
1141	1.24	1.73	60.78	95.77
1154	0.16	0.81	2.99	0.96
1762	1.71	2.21	0.82	7.80
3092	0.79	1.39	1.11	2.36
19	144.25	117.34	0.03	0.02
78	4.80	4.97	0.03	0.08
241	277.40	277.45	0.84	0.20
630	4.56	4.30	0.02	0.28
654	2.69	2.85	0.34	19.83
808	8.43	8.67	0.83	1.34
817	11.90	12.37	1.04	8.51
823	10.30	8.25	1.10	0.03
873	10.43	11.35	0.07	0.03
937	1.18	1.45	10.85	9.96
965	4.51	4.91	23.49	31.39

Table II. Comparison of the lensing Bayes factor $\mathcal{B}_{\text{lens}}$ computed with `DINGO-lensing` and `bilby` for nonlensed and lensed simulated event with different sampling efficiencies ϵ. We report both the sampling efficiency obtained with the lensed (ϵ_{lens}) and nonlensed ($\epsilon_{\text{nonlens}}$) networks.

where θ_1 and θ_1 are the parameters for which the distance is computed, and C_1 and C_2 are their covariances. In this case, $\theta_1 = \{\Delta t, \mu_{\text{rel}}\}$ and $\theta_2 = \{0, 0\}$, defining

$$\mathcal{D}_{\text{nonlens}} = D(\{\Delta t, \mu_{\text{rel}}\}, 0)\,. \tag{C2}$$

Since the prior in $\{\Delta t, \mu_{\text{rel}}\}$ is bounded at 0, the lensing distance satisfies $\mathcal{D}_{\text{nonlens}} \gtrsim 1$. It is to be noted that for non-Gaussian posteriors, the Gaussian approximation will typically lead to smaller distances. Therefore, $\mathcal{D}_{\text{nonlens}}$ is a conservative metric to find lensed candidates. The advantage of $\mathcal{D}_{\text{nonlens}}$ over the lensing Bayes factor $\mathcal{B}_{\text{lens}}$ is that it does not require importance sampling and can be obtained directly from neural posterior estimation with `DINGO-lensing`.

ID	$\mathcal{M}_c/M_\odot$	q	a_1	a_2	ϕ_1	ϕ_2	ϕ_{12}	ϕ_{JL}	d_L [Mpc]	θ_{JN}	ψ	ϕ_{ref}	t_{ref} [s]	RA	DEC	Δt [s]	μ_{rel}
6	31.32	0.43	0.05	0.15	2.18	2.31	3.26	2.15	2098.68	0.36	2.26	0.69	99.92	2.73	-0.60	0.00	0.00
594	34.16	0.42	0.43	0.34	0.83	2.29	3.64	4.25	6192.94	0.59	1.09	4.67	100.10	3.79	0.99	0.00	0.00
779	35.06	0.57	0.10	0.44	1.58	1.73	0.01	4.60	4038.44	1.63	1.60	3.22	99.99	3.87	-1.10	0.00	0.00
1141	52.37	0.39	0.04	0.11	1.69	0.33	4.46	5.86	3538.84	1.01	0.10	4.64	99.96	1.73	0.84	0.00	0.00
1154	33.58	0.96	0.48	0.10	1.12	1.13	5.87	2.30	10810.94	2.78	2.38	2.82	99.99	0.48	1.20	0.00	0.00
1762	32.98	0.72	0.46	0.36	1.13	0.71	0.60	3.74	3771.03	1.91	2.90	4.56	100.04	2.16	1.40	0.00	0.00
3092	37.02	0.42	0.01	0.00	0.89	1.69	0.46	2.25	2125.64	2.35	0.06	0.12	99.99	1.49	-0.27	0.00	0.00
19	37.95	0.40	0.35	0.08	0.23	2.00	4.52	4.52	3608.73	2.41	1.32	0.40	100.02	3.19	-0.84	0.07	0.94
78	37.14	0.79	0.23	0.35	1.99	0.35	3.75	0.27	14037.69	2.68	0.39	1.55	99.90	0.07	1.04	0.04	0.44
241	51.96	0.73	0.34	0.17	0.84	1.25	0.14	4.73	2047.32	0.62	2.80	0.26	99.99	5.84	0.69	0.10	0.30
630	33.38	0.83	0.42	0.36	1.79	2.20	4.92	6.07	8883.59	2.87	2.04	5.37	100.04	4.31	1.05	0.03	0.84
654	43.40	0.86	0.16	0.04	0.72	1.53	5.24	4.67	6622.99	0.97	1.25	1.90	99.91	3.08	1.49	0.09	0.52
808	36.66	0.52	0.17	0.47	0.58	1.48	2.26	1.31	9722.49	2.78	0.79	6.03	100.03	0.01	0.47	0.02	0.81
817	32.48	0.68	0.44	0.39	1.33	2.10	3.87	3.56	3721.55	1.36	1.88	2.22	100.09	1.47	-1.13	0.04	0.52
823	34.49	0.64	0.45	0.06	0.24	1.53	2.45	5.01	6490.86	2.82	1.61	3.77	99.96	4.20	0.77	0.01	0.94
873	30.34	0.52	0.43	0.13	1.25	0.84	5.48	2.06	1710.62	1.71	2.36	3.37	100.07	3.22	-0.25	0.02	0.73
937	43.38	0.63	0.37	0.50	0.77	1.79	4.01	1.48	6277.43	1.39	2.66	0.28	100.06	5.93	0.29	0.03	0.30
965	53.73	0.46	0.24	0.28	1.02	1.30	2.29	2.62	8872.92	0.65	2.77	4.32	99.94	3.41	-0.55	0.08	0.48

Table III. Parameters for the simulated events in Table II.