

Accelerated inference of microlensed gravitational waves with machine learning

Marienza Caldarola^{1,*}, Srashti Goyal^{2,†}, Nihar Gupte^{2,‡}, Stephen R. Green^{3,§} and Miguel Zumalacárregui^{2,||}

¹*Instituto de Física Teórica UAM-CSIC, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain*

²*Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, D-14476 Potsdam-Golm, Germany*

³*School of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD, United Kingdom*



(Received 28 November 2025; revised 26 February 2026; accepted 27 April 2026; published 29 May 2026)

Gravitational waves (GWs) within the LIGO-Virgo-KAGRA sensitivity band can be microlensed by stellar and intermediate-mass black holes, producing a frequency-dependent modulation of the signal amplitude. Microlensing analyses, however, are costly due to the increased dimensionality of the parameter space and waveform computation time. As a proof of concept, we show that the deep-learning-based framework, Deep Inference for Gravitational-Wave Observations (DINGO), which employs a simulation-based inference approach to estimate posterior distributions, can perform efficient parameter inference for GW microlensing by an isolated point-mass lens. Using simulated microlensed GW signals, we train a lensed-DINGO network and compare its performance with traditional Bayesian parameter estimation carried out with Bilby. Our framework can be used to rapidly identify microlensed events in large GW catalogs. When the lensed-DINGO network is combined with importance sampling, we find that, although sample efficiencies are somewhat reduced compared to the unlensed-DINGO network, owing to the richer structure of microlensed signals, it still achieves $\mathcal{O}(10\times)$ speed-up relative to Bilby. We further show that this framework is useful to efficiently estimate the background Bayes-factor distribution, which is crucial for assessing the significance of candidate lensed events. However, for foreground (lensed) events, the sampling efficiency can sometimes drop when analyzed with the unlensed-DINGO network, providing a diagnostic indicator of out-of-distribution data. Our approach can be straightforwardly generalized to more complex and realistic lens models, enabling detailed studies of microlensed GWs.

DOI: [10.1103/lvwm-d7fm](https://doi.org/10.1103/lvwm-d7fm)

I. INTRODUCTION

Gravitational waves (GWs) are ripples in spacetime generated by the acceleration of massive compact objects, such as merging black holes or neutron stars. Since their first detection by LIGO and Virgo Collaboration in 2015 [1], they have become a powerful probe of gravity, astrophysics, and cosmology [2–8].

Like any other signal, GWs traveling through the Universe are deflected and distorted by the gravitational

fields along their path [9]. Gravitational lensing offers a powerful probe of cosmology, astrophysics, and fundamental physics, with applications such as constraining the distribution of dark matter [10–12]. Weak lensing produces small distortions that trace the large-scale matter distribution, while strong lensing constrains the mass profiles of galaxies and clusters [13–15]. Lensing phenomena can enable precision tests of general relativity [16–19] and cosmological parameters [20–22]. Gravitational lensing phenomena are essential to correctly interpret observations across the electromagnetic spectrum [23]. For GWs, in addition to being magnified and occasionally split into multiple images, lensing can produce novel effects due to the signal’s finite wavelength, such as diffraction and interference [24,25]. Accurately modeling these effects is crucial: if not accounted for, lensing can bias the posteriors and introduce systematic errors in source and cosmological inference. Most importantly, diffraction may be the smoking gun for small-scale objects and distant magnified GWs [23,26–35].

GW lensing can be described in two regimes: the geometric optics (GO) regime, where the GW wavelength

*Contact author: marienza.caldarola@csic.es

†Contact author: srashti.goyal@aei.mpg.de

‡Contact author: nihar.gupte@aei.mpg.de

§Contact author: stephen.green2@nottingham.ac.uk

||Contact author: miguel.zumalacarregui@aei.mpg.de

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI. Open access publication funded by the Max Planck Society.

is much smaller than the characteristic gravitational radius of the lens, and the wave-optics (WO) regime, where the GW wavelength is comparable or bigger than the characteristic size of the lens, where diffraction and interference become significant. We focus on the WO regime, as its characteristic interference patterns imprint distinctive, frequency-dependent modulations on the waveform, as clear signatures of lensing by compact objects. These effects are often combined, with a macroscopic lens (a galaxy or cluster) magnifying the GW and small-scale objects (stars, remnants) diffracting it; these cases are referred to as macrolensing and microlensing, respectively [31].

Lensing computations are challenging in the WO regime, as they require solving highly oscillatory integrals [36,37]. Recent advances in algorithms [38–40] and parameter estimation (PE) [41,42] now enable the search for these signatures among GW events. Despite this progress, inference becomes more costly due to the computational cost of diffraction integrals and the extended parameter space. Intriguing candidates for lensed GWs have been reported [43–45], although none yet represent an unambiguous detection [46–48]. Nevertheless, lensed GWs are inevitable due to the rapidly increasing number of observed events, warranting a search among the $\mathcal{O}(100/\text{yr})$ in the current run, the fourth observing run (O4), which will reach $\mathcal{O}(10^5/\text{yr})$ with next-generation detectors [49]. Many forecasts agree that lensed GWs will be commonplace in the next observing run O5 [50–53], but they have to be identified among thousands of signals observed every year.

Given the inevitability of lensed signals within the growing data volume, the challenges of WO computations, and the complexity of traditional analysis, we are motivated to develop a machine learning framework for fast and accurate PE of lensed GW signals. Traditional Bayesian inference methods, such as nested sampling [54–56], are computationally expensive and often prohibitively time consuming for large datasets. Deep learning techniques have emerged as a powerful alternative, enabling rapid and efficient PE in GWs analyses [57,58]. Recent studies, have also explored machine-learning-based inference for GW lensing parameters using various methodologies, pursuing similar objectives to this work. Examples include the use of conditional variational autoencoders [59], normalizing flows for stochastic lensing [35]. While preliminary, our approach demonstrates the potential for extending machine-learning-based microlensing inference to real data in future GW analyses.

In this study, we model lensing effects on GW signals by computing the amplification factor using Gravitational Lensing of Waves (GLoW) [40], a precise and flexible code that calculates frequency-dependent amplification for general lens configurations. PE is then performed using the Deep Inference for Gravitational-Wave Observations (DINGO) framework [60–63]. DINGO has been reviewed by the LIGO-Virgo-KAGRA (LVK) Collaboration, used for eccentricity estimates [64], and extended to work on binary

neutron star systems [65]. Recently, it has also been applied to overlapping GW signals [66].

The paper is organized as follows. In Sec. II, we introduce the gravitational lensing of GWs. In Sec. III, we address the Bayesian PE methods, while in Sec. IV we review the DINGO framework and mention our train setup in Sec. V. In Sec. VI, we present our results and provide conclusions in Sec. VII.

II. GW LENSING

Gravitational lensing in the weak field limit modulates the GW strain by a frequency-dependent amplification factor [9],

$$F(f) \equiv \frac{\tilde{h}_L(f)}{\tilde{h}_0(f)}, \quad (1)$$

where $\tilde{h}_0(f)$ and $\tilde{h}_L(f)$ denote the frequency-domain unlensed and lensed strains, respectively. The amplification factor is calculated from the diffraction integral over the lens plane $\mathbf{x}$,

$$F(w) = \frac{w}{2\pi i} \int d^2x \exp(iw(\phi(\mathbf{x}, \mathbf{y}) - \phi_0)), \quad (2)$$

where $\phi(\mathbf{x}, \mathbf{y}) = \frac{1}{2}|\mathbf{x} - \mathbf{y}|^2 - \psi(\mathbf{x})$ is the Fermat potential, ϕ_0 is its minimum, and $\mathbf{y}$ is the source position, rescaled by an arbitrary length scale ξ_0 . The lensing potential $\psi(\mathbf{x})$ depends on the projected mass density of the lens (through a Poisson equation) and the dimensionless frequency,

$$w \equiv 8\pi G M_{Lz} f \approx 1.24 \left(\frac{M_{Lz}}{100 M_\odot} \right) \left(\frac{f}{100 \text{ Hz}} \right), \quad (3)$$

is a scaled version of GW signal frequency f using the effective lens mass $M_{Lz} \equiv \frac{\xi_0^2(1+z_L)D_S}{4GD_L D_{LS}}$, where D_S is the angular distances to the source, D_L to the lens, and D_{LS} between lens and source. We have set $c = 1$.

In the GO limit ($w \gg 1$), the amplification factor gets a contribution only from the stationary points of the Fermat potential, where the images are formed. However, in the WO regime ($w \sim 1$), the full diffraction integral needs to be evaluated, which for most types of lenses cannot be done analytically. A set of numerical algorithms is packaged in GLoW to evaluate this integral efficiently for various lens models [40].¹ The diffracted signal thus encodes the information about the lens properties and allows one to probe dark and baryonic matter in the Universe [67–73].

Many models are used to describe the mass distribution of lenses [74]. In the simplest case, point lens models, representing compact objects such as planets, stars, or black

¹The GLoW code is publicly available at https://github.com/miguelzuma/GLoW_public

holes, allow for analytical treatment of the lensing properties [75]. More complex astrophysical lenses, such as galaxies or clusters, are described using extended profiles like the Singular Isothermal Sphere [76] and Navarro-Frenk-White [77]. Even then, realistic representations require the superposition of many such simple lenses, often including additional deformations (ellipticity, multipolar structure). The $F(w)$ computations become much more expensive for extended lenses, and machine learning can be particularly helpful in such cases.

As a proof of principle, we will consider only the isolated point-mass lens, neglecting an environmental potential and additional objects. The point lens lensing potential takes the form $\psi(x) = \ln x$. Conveniently, choosing the scale $\xi_0 = R_E$, where $R_E \equiv \sqrt{\frac{4GM_L D_L D_{LS}}{D_S}} \approx 0.14 \text{ pc} \left(\frac{M_L}{100 M_\odot}\right)^{1/2} \left(\frac{D_L D_{LS} D_S^{-1}}{1 \text{ Gpc}}\right)^{1/2}$ as the Einstein radius of the point-mass lens, yields $M_{Lz} = M_L(1 + z_L)$, where M_L is the mass of the lens and z_L is its redshift. M_{Lz} is equivalent to the “redshifted lens mass.” Effectively, the diffraction by a point-mass lens on the GW waveform is given by a frequency-dependent amplification factor $F(f)$ that depends on two parameters: M_{Lz} and y , where y is the impact parameter of the source in units of the Einstein radius. It can be computed analytically as a hypergeometric series expansion, for which several approximations speed up the computation in different regions of the parameter space (Sec. VA in [40]).

In the GO limit, which holds for $w \gg 1$, the lensed waveform is approximately equivalent to an interference of two images. The amplification factor can be written in GO approximation as

$$F(f) \approx \sqrt{\mu_+} + \sqrt{\mu_-} \exp(2\pi i f \Delta t - \pi/2). \quad (4)$$

This gives rise to two images with magnifications $\mu_{\pm} = \frac{1}{2} \left| \frac{y^2 + 2}{y\sqrt{y^2 + 4}} \pm 1 \right|$ (defined positive) and separated by a time delay,

$$\Delta t = 4GM_{Lz} \left[\frac{y}{2} \sqrt{y^2 + 4} + \ln \left(\frac{\sqrt{y^2 + 4} + y}{\sqrt{y^2 + 4} - y} \right) \right] \approx 4 \text{ ms} \left(\frac{M_{Lz}}{1000 M_\odot} \right), \quad \text{at } y = 1. \quad (5)$$

An example of $F(w)$ for a point lens, along with the GO approximation, using different impact parameters y is given in Fig. 1. Notice that, for high frequencies $w \gg 1$ and/or lens masses, the $F(w)$ is very oscillatory; this is where GO is a good approximation. On the other hand, for $w \sim 1$ the GO approximation fails and one needs to consider the full WO amplification factor. As expected, the smaller impact parameters have a more pronounced effect on the waveform than the larger impact parameters. For example, for $y \sim 5$, the lensing amplification is negligible with $F(w) \sim 1$, irrespective of the mass of the lens. The LIGO-Virgo

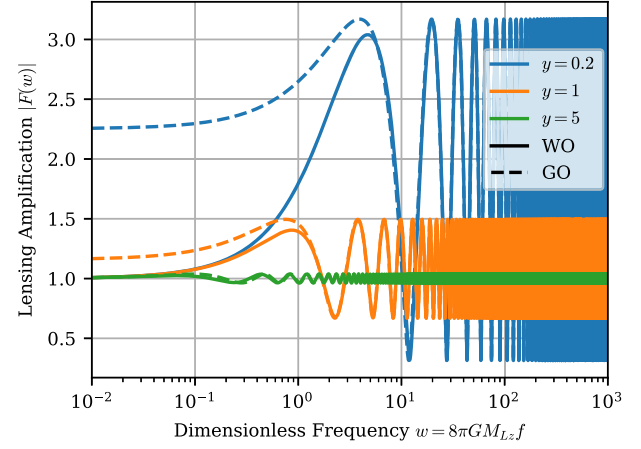


FIG. 1. Lensing amplification factor $F(w)$ for a point-mass lens at different impact parameters y , in units of the Einstein radius R_E . The WO regime is near $w \sim 1$, whereas for $w \gg 1$ the GO approximation is valid. The smaller impact parameters modulate the waveform much more than the larger impact parameters, such that for $y \gg 1$ the lensing effect is negligible $F(w) \sim 1$.

detectors are most sensitive in the frequency range $f \in [10, 1000]$ Hz, which corresponds to the WO lensing regime for isolated point masses with $M_{Lz} \in [10, 10^4] M_\odot$.

III. BAYESIAN PE

In the context of GW analysis, the observed detector data are given by $d(t) = h(t, \theta) + n(t)$, where $h(t, \theta)$ is a GW strain template for the waveform given the source parameters θ and $n(t)$ the detector noise. The Bayes theorem is used to infer the properties of an observed GW source,

$$p(\theta|d, H) = \frac{p(d|\theta, H)p(\theta|H)}{p(d|H)}, \quad (6)$$

where $p(\theta|d, H)$ is the posterior probability distribution of the parameters θ given the data d and the hypothesis H , $p(\theta|H)$ is the prior, and $p(d|\theta, H)$ is the likelihood. The denominator $p(d|H)$ is Bayesian evidence Z_H , which is useful for model selection. Given two hypotheses, in our case lensed H_L and unlensed H_U , the Bayes factor is the ratio of evidences $B_U^L = Z_L/Z_U = P(d|H_L)/P(d|H_U)$. This ratio quantifies the preference of one hypothesis over the other, considering only the data.

In GW astronomy, a Gaussian noise likelihood function is generally assumed, which in the frequency domain takes the form [78]

$$p(d|\theta) \propto \exp \left(-2 \int_0^\infty df \frac{|\tilde{d}(f) - \tilde{h}(f, \theta)|^2}{S_n(f)} \right), \quad (7)$$

where $\tilde{h}(f, \theta)$ represents the GW signal in the frequency domain for a given set of parameters θ , projected onto the detector antenna response pattern, while S_n is the power

spectral density (PSD) of the detector noise. For any signal $\tilde{h}(f)$, its optimal signal-to-noise ratio (SNR) is given by $\rho_{\text{opt}} = [4 \int_0^\infty \frac{|\tilde{h}(f)|^2}{S_n(f)} df]^{1/2}$.

To estimate the posterior, one typically relies on evaluating the likelihood explicitly for many parameter samples from the prior while using efficient sampling techniques like nested sampling [54]. In GW astronomy, often one uses the Bilby [56] code, which is a GW specific interface to many nested sampling implementations. The most popular of these is DYNESTY [55].

Given the 15-dimensional parameter space for a binary black hole, and an additional two parameters in the case of the isolated point lens, sampling turns out to be computationally expensive taking $\mathcal{O}(\text{days})$. It is especially challenging for constructing a background distribution of Bayes factors, which requires a large number of unlensed injections (e.g., currently ~ 400 injections for 3σ deep background) to estimate the significance of a lensed event candidate. These methods are not feasible given the expected number of GW events, which will rise as detector sensitivities improve, reaching $\mathcal{O}(10^5)$ binary black hole (BBH) events per year for next-generation detectors [79]. This motivated us to use the simulation-based inference tool DINGO [60,61].

IV. DINGO FRAMEWORK

DINGO² utilizes neural posterior estimation (NPE) with normalizing flows to model the posterior distribution $p(\theta|d)$. The flow f_ϕ learns an invertible transformation from a simple base distribution $p(\mathbf{u})$ to the complex posterior, conditioned on the data: $\theta = f_\phi(\mathbf{u}; d)$ [80]. The resulting probability density, $q_\phi(\theta|d)$, is given by the change of variables formula,

$$q_\phi(\theta|d) = p(\mathbf{u} = f_\phi^{-1}(\theta; d)) \left| \det J_{f_\phi^{-1}}(\theta; d) \right|, \quad (8)$$

where $J_{f_\phi^{-1}}$ is the Jacobian of the inverse transformation. The network is trained by minimizing the forward Kullback-Leibler divergence between its output distribution $q_\phi(\theta|d)$ and the true posterior $p(\theta|d)$. This is achieved by minimizing the negative log-likelihood on a training set of simulated signal-parameter pairs (θ_i, d_i) , generated by injecting GW signals into stationary Gaussian noise realizations. The loss function is the negative log-likelihood, averaged over the joint distribution of parameters and data, $p(\theta, d)$,

$$\mathcal{L}(\phi) = \mathbb{E}_{p(\theta, d)} [-\log q_\phi(\theta|d)]. \quad (9)$$

Since the true posterior is unknown, this loss is expressed as an expectation over the prior $p(\theta)$ and the likelihood

$p(d|\theta)$, which can be approximated using a Monte Carlo average over training samples,

$$\mathcal{L}(\phi) \approx -\frac{1}{N} \sum_{\substack{\theta^{(i)} \sim p(\theta) \\ d^{(i)} \sim p(d|\theta^{(i)})}} \log q_\phi(\theta^{(i)}|d^{(i)}), \quad (10)$$

where N is the number of samples in the minibatch. This training process allows the network to perform amortized inference, generating posterior samples for new events in seconds, without evaluating any likelihoods explicitly. This approach is called “simulation-based inference.”

To simplify the learning task for the primary NPE network, we infer the coalescence time in each detector using the group-equivariant neural posterior estimation (GNPE) algorithm [60,81]. This approach standardizes the time of arrival of the signal across the detector network. Although this requires training a separate neural network, it simplifies the data by aligning coalescence times, allowing the main NPE network to focus more effectively on the remaining physical parameters of the source.

DINGO also allows for importance sampling (IS), in which the output samples are corrected by explicit likelihood evaluation [63]. Given our target distribution $p(\theta|d) \propto p(d|\theta)p(\theta)$ and a set of n samples from the proposal distribution, each sample $\theta_i \sim q(\theta|d)$ is assigned an importance weight,

$$w_i = \frac{p(d|\theta_i)p(\theta_i)}{q(\theta_i|d)}. \quad (11)$$

Since the proposal is unlikely to be perfect, the effective sample size

$$n_{\text{eff}} = \frac{(\sum_i w_i)^2}{\sum_i w_i^2} < n, \quad (12)$$

resulting in a sample efficiency

$$\epsilon = \frac{n_{\text{eff}}}{n}. \quad (13)$$

The sample efficiency provides a measure of proposal quality [63]. IS requires evaluating $p(d|\theta)p(\theta)$ for all n samples. For sufficiently large ϵ , DINGO IS requires many fewer likelihood evaluations than conventional sampling techniques, and moreover, they can be executed in parallel. The evidence $Z = p(d) = \int d\theta P(d|\theta)P(\theta)$ is estimated from the normalization of the weights,

$$Z = p(d) = \frac{1}{n} \sum_i w_i. \quad (14)$$

IS corrects for any mismatch between the proposal distribution and the true posterior and thereby improves the final inference results. The statistical uncertainty of the IS evidence estimator is controlled by the variance of the importance weights [63],

²<https://github.com/dingo-gw/dingo>

TABLE I. Priors used for preparing the datasets for training DINGO. The bottom two rows indicate the lens parameters prior used to generate training dataset for lensed DINGO. The same priors are used for the Bilby.

Symbol	Parameter name	Prior (range)
$\mathcal{M}_c$	Chirp mass	Uniform [15, 150] $M_\odot$
q	Mass ratio	Uniform [0.125, 1.0]
a_1	Primary spin magnitude	Uniform [0, 0.99]
a_2	Secondary spin magnitude	Uniform [0, 0.99]
θ_1	Primary spin tilt angle	Sine [0, π]
θ_2	Secondary spin tilt angle	Sine [0, π]
ϕ_{12}	Spin azimuthal angle difference	Uniform [0, 2π]
ϕ_{jl}	Precession phase	Uniform [0, 2π]
t_c	Coalescence time	Uniform [-0.1, 0.1] s
d_L	Luminosity distance	Uniform [100, 3000] Mpc
θ	Inclination angle	Sine [0, π]
ϕ_c	Coalescence phase	Uniform [0, 2π]
α	Right ascension	Uniform [0, 2π]
δ	Declination	Cosine [$-\pi/2$, $\pi/2$]
ψ	Polarization angle	Uniform [0, π]
y	Impact parameter	Uniform [0.1, 5]
$\log_{10} M_{Lz}$	Redshifted lens mass ($\log_{10}$)	Uniform [1, 4] $M_\odot$

$$\sigma^2(Z) = \frac{1}{n} \sigma^2(w) = \frac{1}{n} (\langle w^2 \rangle - \langle w \rangle^2). \quad (15)$$

Consequently, the relative error on the evidence scales as

$$\frac{\sigma(Z)}{Z} \propto \sqrt{\frac{1}{n_{\text{eff}}}}. \quad (16)$$

Hence, for an accurate evidence estimate, the effective sample size should be large. Meaning, for lower sampling efficiency, one would need a larger number of proposal samples (n) to achieve the desired n_{eff} . In all cases considered in this study, a conservative requirement on the effective sample size is imposed ($n_{\text{eff}} \gtrsim 500$), ensuring that the resulting evidence estimates and Bayes factors are not dominated by a small number of high-weight samples.

V. TRAINING PROCESS

For training lensed-DINGO network, we implement a WO lensing transform that multiplies $F(f)$ (see Sec. II) to simulated signals on the fly during training. This maximizes coverage of the lensing parameters and mitigates overfitting. As the point lens waveform computation is very fast ($\mathcal{O}(0.1 \text{ ms})$), this is a reasonable approach; however, for more complicated lensed models, lensed waveforms should be computed in advance of training.³ We evaluate the amplification factor on the signals using GLoW for a point lens in the WO regime (see Sec. II).

³All modifications to include lensing in DINGO have been developed in a branch, available publicly at <https://github.com/sra-shtig/dingo/tree/lensing>

In addition to lensed signals, we also train DINGO without any lensing transform, representing the standard analysis (15 parameters), to compare the results. From now on, we will distinguish the two networks as unlensed DINGO and lensed DINGO. For both networks, we used 10^7 waveforms in the frequency domain based on the IMRPhenomXPHM approximant [82], sampled at 2048 Hz, in the range [20, 1024] Hz with a duration of 8 sec. The GW signals are injected into the Gaussian noise, sampled from a fixed PSD of the Hanford, Livingston, and Virgo detectors using the data from the third observing run (O3) [83]. We estimate PSDs using the Welch method with 1024 sec of data and a sampling frequency of 4096 Hz. The training dataset is generated using the standard prior distributions for usual 15 parameters of the source, see Table I. The lens parameters are generated for the redshifted lens mass, $\log_{10}(M_{Lz}) \in \text{Uniform}[1.0, 4.0]$, and the impact parameter, $y \in \text{Uniform}[0.1, 5.0]$, as shown in Table I.

We train networks using one 40 GB NVIDIA A100 GPU, with 16 CPUs for data preprocessing. We use a batch size of 4096 and the network architecture settings from [60]. We train both networks for 275 epochs, which takes approximately two weeks. The loss functions for both unlensed DINGO and lensed DINGO are reported in Fig. 2.

The main computational cost lies in the training of the networks, which requires approximately two weeks to fully train the model. However, once trained, drawing 10^5 samples from the network takes just a few minutes.

VI. TEST RESULTS

In this section we begin by performing a statistical test called a “probability-probability” ($p - p$) test for network

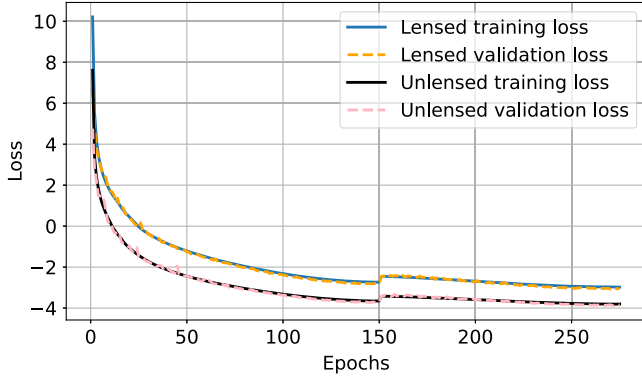


FIG. 2. Evolution of the training and validation loss for both unlensed DINGO and lensed DINGO as a function of epochs. It demonstrates similar training and validation losses, indicating no overfitting. Note that one cannot compare directly the losses of the two networks as they have different priors and number of parameters.

validation. Then we compare the results with Bayesian PE using Bilby. Finally, we investigate why lensing inference is challenging compared to the unlensed case.

A. $p-p$ test

To validate the lensed-DINGO network, we conduct a $p-p$ test using 1000 lensed injections drawn from the prior in Table I, each recovered using lensed DINGO

without IS. For each injection, we determine the p -credible region in which the true parameter lies within the corresponding posterior. If the inference is well calibrated, these cumulative probabilities, over many injections, should follow a uniform distribution between 0 and 1.

Figure 3 displays the resulting $p-p$ plots for the lensed-DINGO network, where all parameters align closely with the diagonal $p = \text{CDF}(p)$, where CDF denotes the cumulative distribution function, confirming that the network remains well calibrated even without IS. We note that the coalescence time t_c exhibits the lowest p value, as the network may be having difficulty inferring the coalescence time due to a degeneracy in lensed waveforms, which are equivalent to the interference between two time-delayed waveforms for higher lens masses. This results in a bimodal t_c posterior sometimes, which gets resolved with increased number of GNPE iterations and after IS. Nevertheless, all other parameters are recovered within 2σ confidence interval, indicating that lensed DINGO can be used effectively, even without IS, for rapid identification of lensed events and for population studies. Note that this test has already been performed for the unlensed-DINGO network in previous studies [60,63].

B. Comparison with Bayesian PE

We compare the results of lensed DINGO and unlensed DINGO with Bayesian PE using Bilby for a simulated black hole binary merger injected into Gaussian noise.

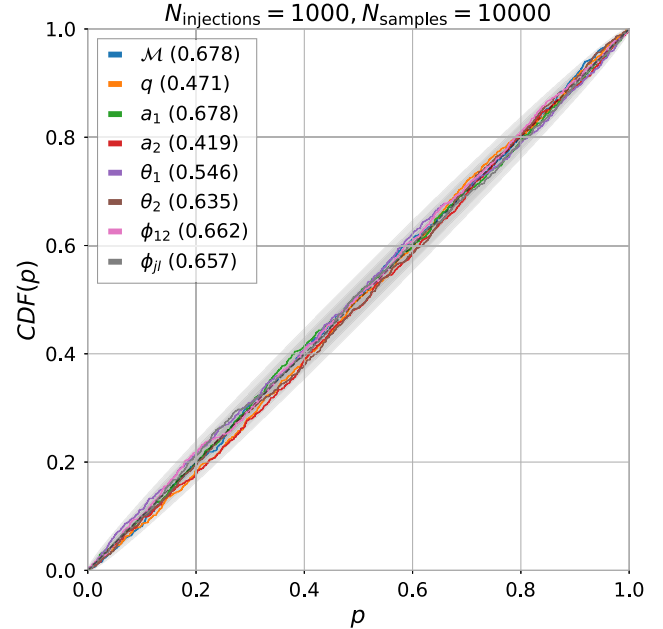
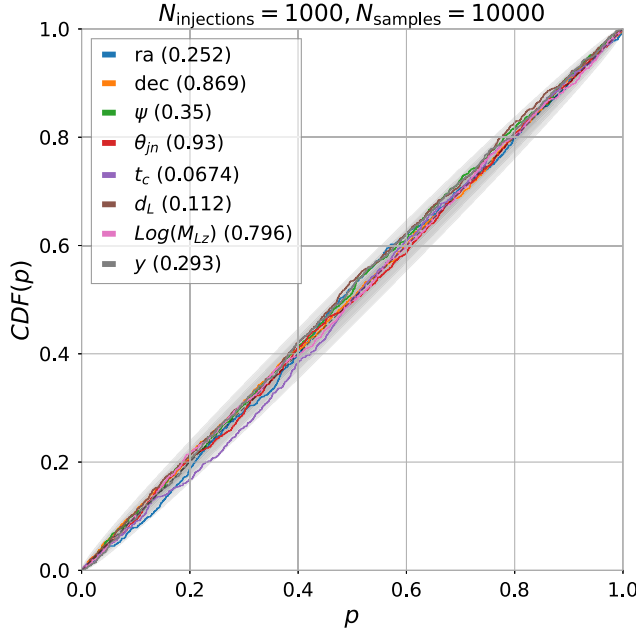


FIG. 3. $p-p$ plots comparing the posterior distributions predicted by DINGO without IS. The numbers in parentheses indicate the corresponding p values associated with each curve, while $N_{\text{injections}}$ and N_{samples} denote the number of simulated injections and the number of posterior samples drawn per injection, respectively. The alignment with the diagonal line indicates a good calibration of the model, that is the inferred parameters fall within the predicted posterior probability regions at the correct rates. As shown by the gray shaded regions, all parameters lie within 2σ confidence interval, except the coalescence time t_c , which is recovered within 3σ . We divided in extrinsic and intrinsic parameters for better visualization.

TABLE II. Parameters corresponding to lensed and unlensed injections used for testing DINGO and comparing with Bilby.

Parameter	Lensed $y = 0.2$ (1.2)	Unlensed
$\mathcal{M}_c$	$28.10 M_\odot$	$28.10 M_\odot$
q	0.8	0.8
a_1	0.4	0.4
a_2	0.3	0.3
θ_1	0.5 rad	0.5 rad
θ_2	1.0 rad	1.0 rad
ϕ_{12}	1.7 rad	1.7 rad
ϕ_{jl}	0.3 rad	0.3 rad
t_c	0 s	0 s
d_L	1640 Mpc	820 Mpc
θ	0.4 rad	0.4 rad
ϕ	1.3 rad	1.3 rad
α	1.375 rad	1.375 rad
δ	1.121 rad	1.121 rad
ψ	2.659 rad	2.659 rad
y	0.2 (1.2)	...
$\log(M_{Lz})$	3.0	...
Optimal SNR	35.7 (18.1)	32

We particularly focus on the following three cases: (i) an unlensed injection (GW150914-like), (ii) lensed injection with $y = 1.2$, $M_{Lz} = 1000 M_\odot$, and (iii) lensed injection with $y = 0.2$ and $M_{Lz} = 1000 M_\odot$. The full parameters for these injections are listed in Table II.

Figure 4 shows the waveform corresponding to these injections. As seen in the figure, the lensed signals are much richer than the corresponding base (unlensed) waveform $\tilde{h}_0(f)$ for smaller impact parameters and, for $y = 5$, the lensing modulations are almost undetectable. Note that the unlensed injection waveform is the same as $\tilde{h}_0(f)$ but with a smaller luminosity distance and hence higher amplitude.

For each of these injections, we perform PE using lensed DINGO, unlensed DINGO, and with Bilby under lensed and

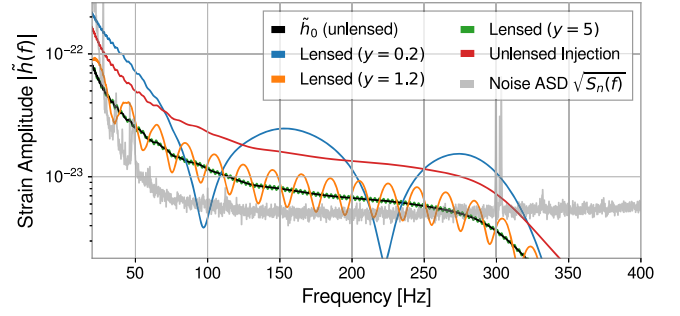


FIG. 4. Waveforms for the injected lensed and unlensed binary black hole merger are shown. The lens is modeled as a point mass with $M_{Lz} = 1000 M_\odot$ and various values of y . Compared to the corresponding unlensed waveform (black), the lensed signals display a more complex modulation pattern because they include the WO modulation factor $F(f)$. The noise amplitude spectral density (ASD = $\sqrt{S_n(f)}$) is plotted in gray. The lensed waveform with $y = 5$ (green) is very similar to the unlensed signal: although the lensing effect is present, the microlensing-induced distortions are too small to be clearly distinguished from an unlensed waveform with current detector network sensitivities. The injection parameters are listed in Table II.

unlensed hypothesis (H_L and H_U). We also perform IS for DINGO posteriors, generating $n = 10^6$ samples to get improved results (DINGO IS), while computing sampling efficiency ϵ , effective sample size n_{eff} , and log evidence $\log Z$. We also compare log evidence and the Bayes factors B_U^L between DINGO IS and Bilby for the three injections, which is summarized in Table III. We now discuss our findings for each of these cases in detail.

- (i) Unlensed injection: Figure 6 shows the posteriors obtained with lensed DINGO and Bilby, which agree closely. As expected, y saturates at its prior maximum ($y = 5$), while $\log_{10}(M_{Lz})$ remains unconstrained. This behavior occurs because, for large impact parameters, the lensed waveform becomes

TABLE III. Comparison of log evidence under lensed and unlensed hypotheses between DINGO IS and Bilby for the three injections. The last two rows compare the Bayes factors $\log B_U^L$. All cases agree well with Bilby, except for the lensed injection with $y = 0.2$ analyzed with unlensed DINGO. In this case, the data are out of distribution for the unlensed network, causing the sampling efficiency to drop to $\epsilon \sim 0.0001\%$ with an effective sample size of $n_{\text{eff}} = 1$, rendering the DINGO Bayes factor unreliable. However, for cases where the effective sample size is sufficient (e.g., $y = 1.2$), DINGO Bayes factors match Bilby well and allow for fast background distribution computation.

	Unlensed	Lensed ($y = 1.2$)	Lensed ($y = 0.2$)
Lensed DINGO	$\log Z = 497.1$ $\epsilon = 0.11\%$, $n_{\text{eff}} = 1110$	$\log Z = 133.9$ $\epsilon = 0.74\%$, $n_{\text{eff}} = 7442$	$\log Z = 592.3$ $\epsilon = 0.06\%$, $n_{\text{eff}} = 602$
Bilby $\log Z_L$	498.9	134.7	593.8
Unlensed DINGO	$\log Z = 497.9$ $\epsilon = 1.8\%$, $n_{\text{eff}} = 17936$	$\log Z = 119.9$ $\epsilon = 10.3\%$, $n_{\text{eff}} = 103233$	$\log Z = 467.4$ $\epsilon = 0.0001\%$, $n_{\text{eff}} = 1$ (unreliable)
Bilby $\log Z_U$	499.7	121.45	482.9
DINGO $\log B_U^L$	-0.8	14.0	124.9
Bilby $\log B_U^L$	-0.8	13.25	110.9

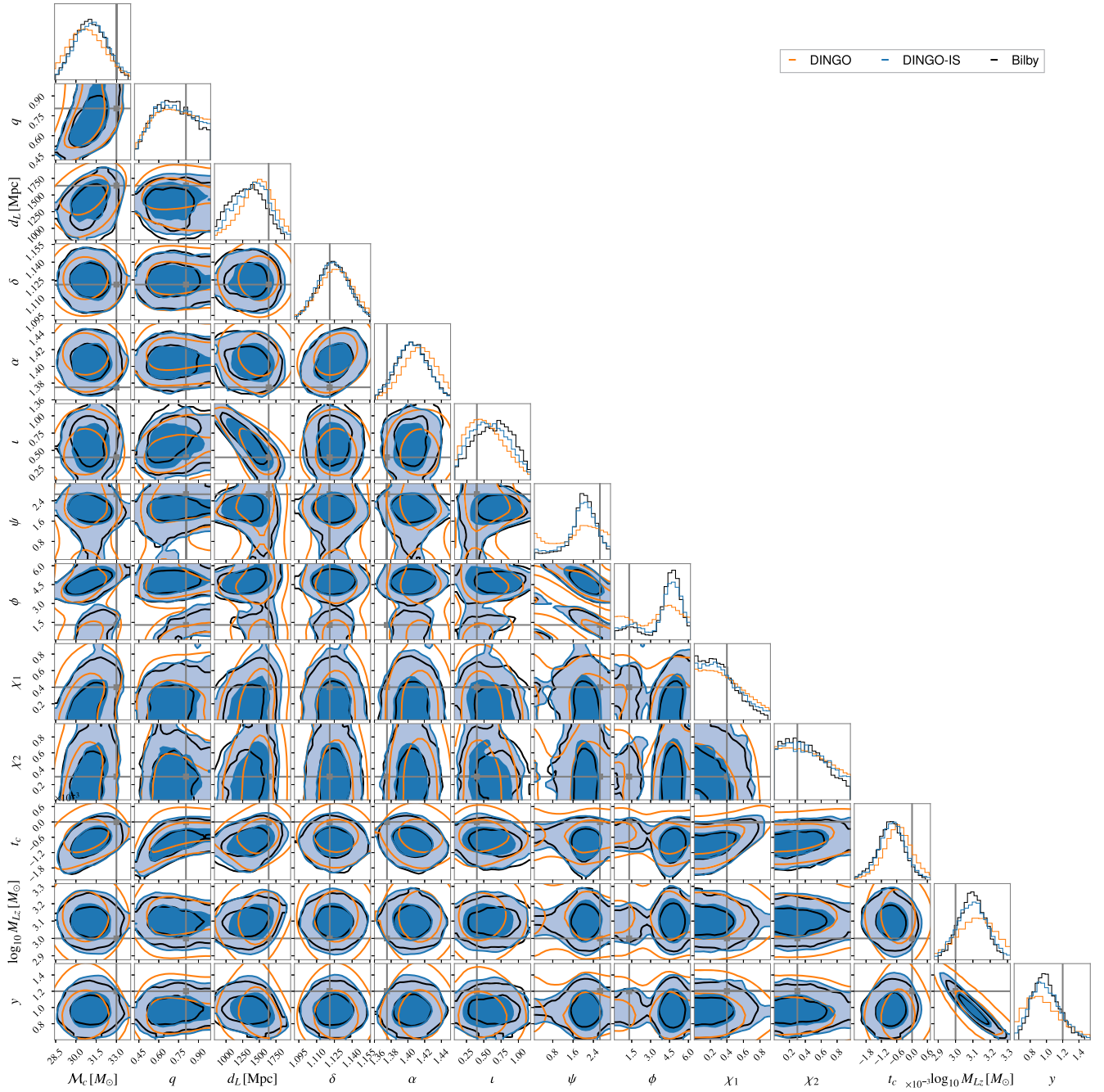


FIG. 5. Lensed ($y = 1.2$) injection with lensed DINGO and Bilby under lensed hypothesis. DINGO posteriors match that of Bilby, especially after IS while recovering the injection (gray). We obtain $\epsilon = 0.74\%$, $n_{\text{eff}} = 7442$, and lensed DINGO $\log Z = 133.9$ which is consistent with Bilby $\log Z_L = 134.7$.

nearly indistinguishable from the unlensed one; in particular, for $y = 5$ the two are indistinguishable, as illustrated in Fig. 4.

Thus, the marginalized posterior probability $p(y = 5|d)$ inferred with lensed DINGO can be used as a fast diagnostic to flag potential lensed events, by imposing a selection criterion of the form $p(y = 5|d) < \text{threshold}$. The effectiveness of this classification strategy can be assessed using a large

set of lensed and unlensed injections, which we defer to future work. The other parameters are recovered consistently with their injected values. In this case, the sampling efficiency is $\epsilon \sim 0.11\%$.

For the same unlensed injection analyzed with unlensed DINGO, we obtain a sampling efficiency of $\epsilon = 1.8\%$, together with good agreement with the Bilby posteriors (see Fig. 1 in Supplemental Material [84]), consistent with previous studies [62,64,81].

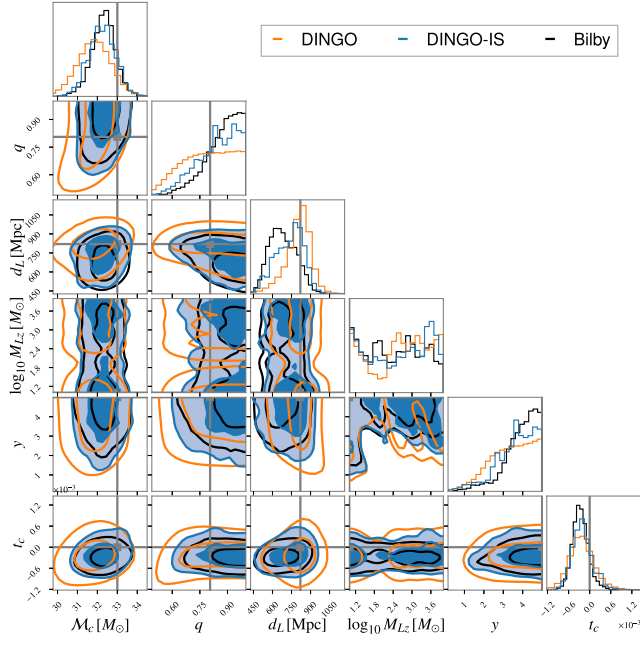


FIG. 6. Unlensed injection with lensed DINGO and Bilby under unlensed hypothesis. Posterior distributions from both DINGO (orange) and DINGO IS (blue) rail at the prior boundary of the impact parameter $y = 5$, and nearly flat in $\log M_{Lz}$ and recover the source parameters (gray). (We do not show spin angles due to size constraints, even though we infer them.) For this injection, we obtain $\epsilon = 0.11\%$, $n_{\text{eff}} = 1110$, and lensed DINGO $\log Z = 497.1$ (consistent with Bilby $\log Z_L = 498.9$).

Since for unlensed injections both lensed DINGO and unlensed DINGO perform well and yield reliable estimates of $\log Z$ and $\log B_U^L$ (see Table III), DINGO is particularly well suited for constructing the background distribution of Bayes factors needed to assess the significance of candidate lensed events, a task for which traditional methods such as Bilby are computationally very expensive.

- (ii) Lensed injection ($y = 1.2$): In Fig. 5, the lensed-DINGO posteriors agree closely with those obtained using Bilby, especially after applying IS, and they recover the injected parameters. The sampling efficiency in this case is $\epsilon \sim 0.74\%$.

The same injection with the unlensed-DINGO network also performs well and agrees with Bilby, achieving a sampling efficiency of $\epsilon \sim 10.3\%$. A subset of its posterior is displayed in Fig. 7(a). Thus, for this particular lensed injection, we can obtain reliable Bayes factors B_U^L . However, as we will show in the next case, this does not always hold.

- (iii) Lensed injection ($y = 0.2$): Here the lensing-induced modulation is stronger than in the previous case. In Fig. 8, the lensed-DINGO posteriors, and therefore the corresponding values of $\log Z$, are consistent with those obtained using Bilby. For this analysis, we measure a sampling efficiency of $\epsilon \sim 0.06\%$

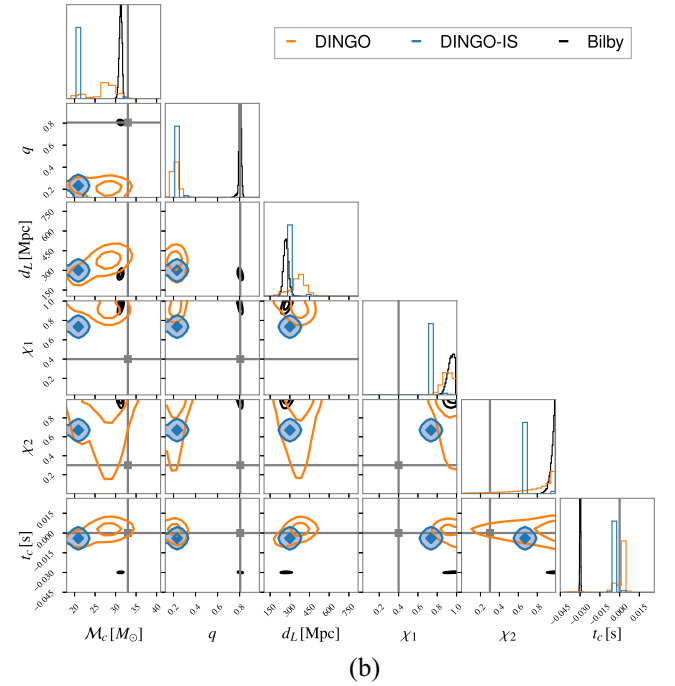
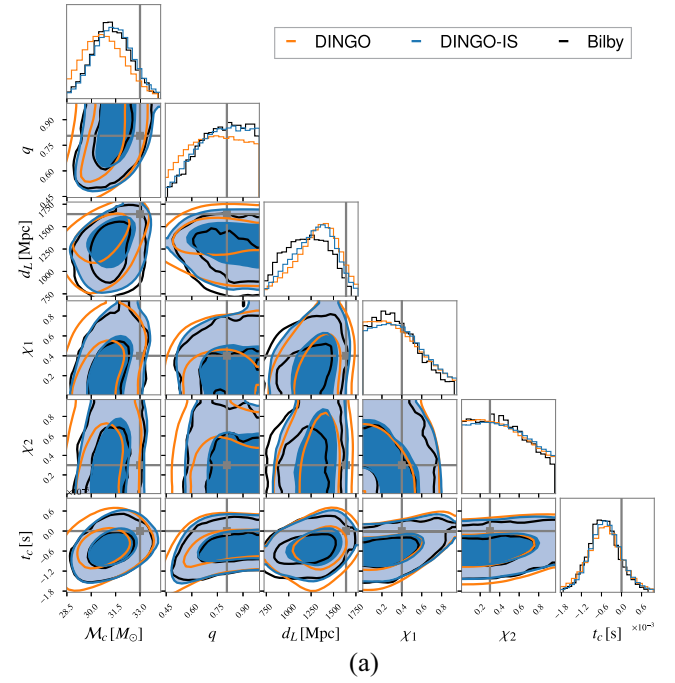


FIG. 7. Lensed injections with unlensed DINGO and Bilby under unlensed hypothesis. (a) For lensed ($y = 1.2$) injection, the unlensed-DINGO network performs well with high sampling efficiency $\epsilon = 10.3\%$ ($n_{\text{eff}} = 103233$), yielding $\log Z = 119.9$ which matches Bilby $\log Z_U = 121.45$. (b) For lensed ($y = 0.2$) injection, the strong lensing features effectively create out-of-distribution data for the unlensed network. The sampling efficiency drops to $\epsilon = 0.0001\%$ ($n_{\text{eff}} = 1$), resulting in unreliable posteriors and evidence estimates ($\log Z = 467.4$ vs Bilby $\log Z_U = 482.9$). However, this low efficiency serves as a diagnostic flag for potential lensing. Notice that here the true (Bilby) posteriors are biased more from the injection (gray) compared to the above case because of the unaccounted microlensing signatures, as expected.

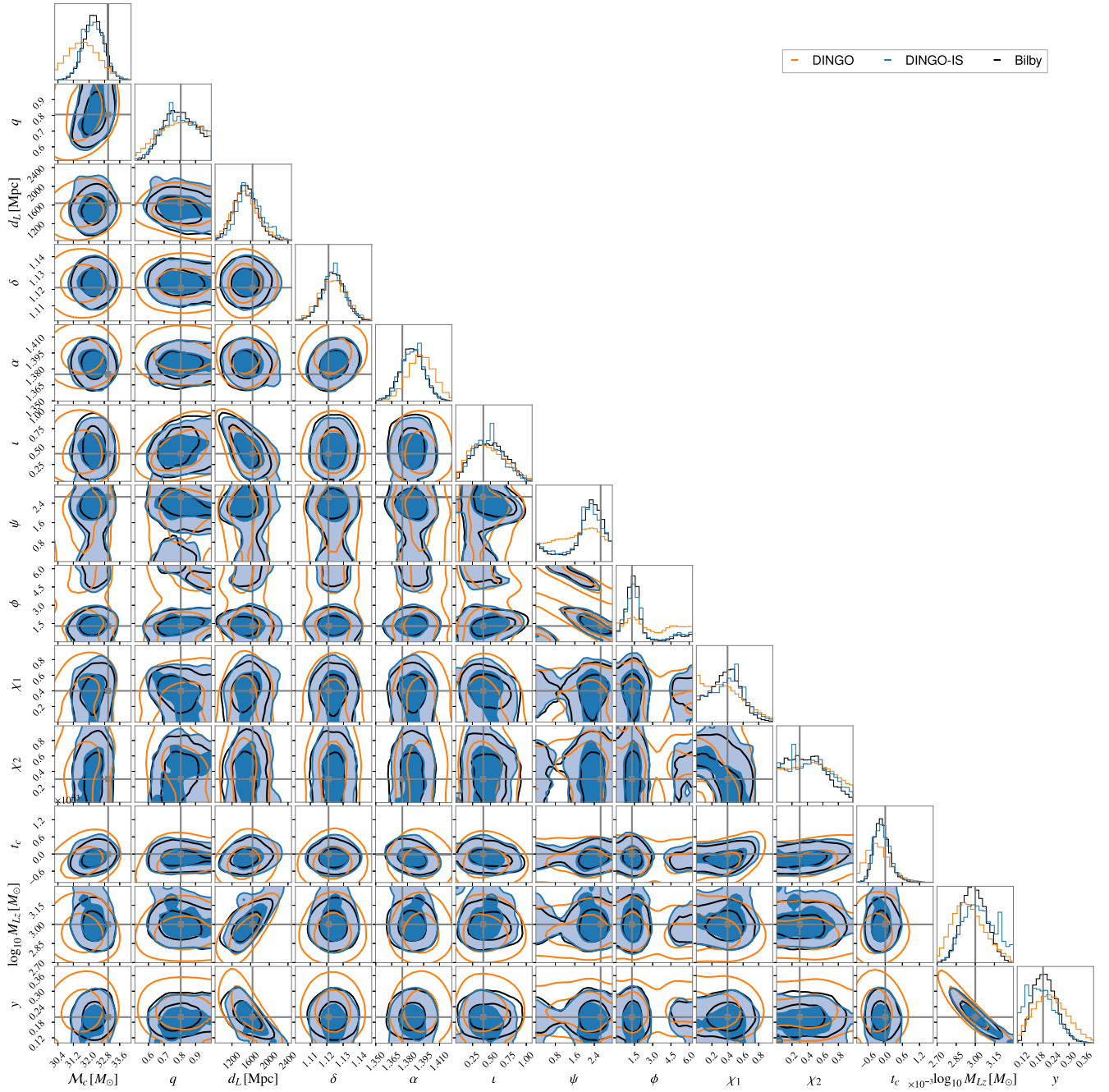


FIG. 8. Lensed ($y = 0.2$) injection with lensed DINGO and Bilby under lensed hypothesis. DINGO posteriors match that of Bilby, especially after IS while recovering the injection (gray). We obtain $\epsilon = 0.06\%$, $n_{\text{eff}} = 602$, and lensed DINGO $\log Z = 592.3$ which is consistent with Bilby $\log Z_L = 593.8$.

In contrast, when we analyze the same injection with unlensed DINGO, the sampling efficiency plummets to $\sim 0.0001\%$, and the effective sample size is reduced to $n_{\text{eff}} = 1$. This behavior arises because the signal is effectively “out of distribution” relative to the network’s training set. A subset of the inferred posteriors, displayed in Fig. 7(b), shows that the DINGO and DINGO-IS posteriors peak at different parameter values, and neither agrees with the Bilby posterior. This demonstrates that, for lensed injections

analyzed under the unlensed hypothesis, both the inferred posteriors and the evidence become unreliable (see Table III) once the sampling efficiency is very low.

Nonetheless, the pronounced drop in efficiency still serves as a useful diagnostic to flag potential out-of-distribution events, such as lensed signals or instrumental artifacts. We emphasize that such an efficiency drop will not always occur, particularly for weakly lensed events or those with low SNR, as in the previous example, where the

unlensed hypothesis did not introduce substantial parameter biases. Consequently, for foreground (lensed) events, one must exercise caution when using DINGO to compute Bayes factors. Reliable identification of candidate lensed events requires that the sampling efficiency and effective sample size be sufficiently high (e.g., $n_{\text{eff}} > 500$).

In summary, we find that after IS, the log-evidence values and posterior distributions for all scenarios (the three injections analyzed with both lensed and unlensed DINGO networks) are in good agreement with Bilby, with one exception: the lensed injection with $y = 0.2$ analyzed using unlensed DINGO, where the pronounced reduction in sampling efficiency clearly indicates out-of-distribution data. We also notice that on an average the sampling efficiencies of lensed DINGO are lower than the unlensed DINGO (see Table III), which can be attributed to the low compressibility of the lensed waveforms. We point the reader to the Appendix for details.

To assess the sampling time between Bilby and lensed DINGO, we can compare the number of likelihood evaluations from each sampler. Given that our sample efficiency ranges from 0.06% to 0.74% in Table III, and if we require 5000 effective samples, we estimate we need 6.8×10^5 – 8.3×10^6 likelihood evaluations per event. This is doubled due to the phase reconstruction step, bringing the total to 1.4×10^6 – 1.7×10^7 evaluations. For the runs in Table III, Bilby computes 6.5×10^7 likelihood evaluations. Therefore, we obtain an $\sim 4\times$ to $\sim 48\times$ [i.e., $\mathcal{O}(10\times)$] speedup in total CPU time. We note that, while this comparison is in terms of CPU time, the likelihood evaluations for DINGO can be parallelized across nodes, further reducing the wall-clock time. We note that DINGO also has the GPU sampling time, which is $\mathcal{O}(h)$ per event for generating 10^6 proposal samples. In practice, for the injections above under lensed hypothesis, the Bilby run took ~ 4.5 days on four CPU cores for us, whereas DINGO IS gave the result in less than 2 h using hyperparallelization. Overall, DINGO IS is substantially faster than Bilby and will be advantageous for efficiently analyzing large numbers of events in forthcoming observing runs. Note that DINGO is expected to remain numerically stable even at SNRs lower than those studied here and can, in fact, more readily capture the broader posteriors that arise in such situations. However, in this regime, microlensing signatures are anticipated to become increasingly indistinguishable from detector noise, resulting in posteriors that are largely dominated by the prior.

VII. CONCLUSIONS

GW lensing provides a novel window into cosmology and astrophysics, enabling tests of dark matter, constraints on cosmological parameters, and insights into the Universe's history. However, extracting lensing signatures can be challenging due to the lens complexity and high computational cost of WO inference. As a proof of concept,

in this work we implemented lensing modifications in DINGO through GLOW and trained lensed-DINGO network to infer parameters for an isolated point-mass lens. The lensed-DINGO network can serve as an efficient alternative for microlensing PE and is well calibrated as per the $p - p$ test, see Fig. 3. It can generate $\sim 10^5$ samples/min (without IS) and enable real-time flagging of potentially microlensed events by monitoring the posterior $p(y|d)$. After IS, one can estimate the evidence and the posteriors are more reliable, consistent with Bilby. Depending on the sampling efficiency, the inference is $\mathcal{O}(10\times)$ faster than Bilby. Hence, lensed DINGO can be utilized to study a large population of microlensed events.

Analyzing microlensed signals with the unlensed-DINGO network may sometimes result in a dramatic drop in sampling efficiency ($\epsilon \sim 10^{-6}$), leading to unreliable posteriors and evidence of Bayes-factor estimates. This serves as a useful diagnostic tool to flag out-of-distribution events, such as microlensed signals or instrumental artifacts. Since both lensed DINGO and unlensed DINGO give reliable posteriors for unlensed events, and are substantially faster than Bilby, they are uniquely suited for computing the background distributions of Bayes factors. This is useful for efficient estimation of the statistical significance of candidate microlensed events across thousands of noise realizations [85]. Note that our lensed-DINGO network achieves comparatively lower sampling efficiencies than unlensed DINGO and those reported in earlier works [64,81], potentially due to the low compressibility of lensed waveforms, see the Appendix. To enhance the performance of lensed DINGO in the future, one could either simplify the input waveforms by calibrating the arrival time of the second microimage using GNPE or increase the complexity of the embedding network [86].

Throughout this study, we considered simulated GW signals embedded in stationary Gaussian noise. Although no microlensing candidates were identified during the third observing run of LVK, future work will focus on extending this analysis to real events. The DINGO framework has already been successfully deployed to analyze real events, and with the methodology developed here, future work will involve reanalysis of potential microlensed candidates, such as GW231123 in the fourth observing run by retraining with updated noise conditions [48,87].

Overall, this study demonstrates that combining machine learning with microlensing models provides an efficient and reliable framework for rapid inference, paving the way for real-time analysis and large-scale population studies of microlensed events. In the future, this can be expanded to include more complex lens models (e.g., lenses embedded in external potentials [87,88]), orbital effects (e.g., eccentricity), and degeneracies with other phenomena (e.g., overlapping signals [89,90]). Additionally, the current DINGO and lensed-DINGO implementations are restricted to scenarios in which the lensed signal can be treated as a

single, stationary data segment in the frequency domain. Strong lensing configurations involving resolvable or partially resolvable time delays between multiple images, such as cases where signals overlap only partially or arrive as distinct events, may require a time-domain treatment [91] and therefore lie beyond the scope of the present framework. Recently, Ref. [92] carried out a related analysis applying DINGO to lensed events, but modeling the lensing through the GO approximation as overlapping signals. Since the GO approximation is appropriate for higher lens masses (and $w \gg 1$, see Fig. 1), whereas we restrict the lens mass to $M_{Lz} = 10^4 M_\odot$, the two studies are complementary.

As the cost of evaluating the WO lensing amplification factor rises in more complex lens systems, machine learning can facilitate in-depth analysis using multiple lens models and waveform approximants.

ACKNOWLEDGMENTS

We are very grateful to Jonathan Gair for valuable discussions and Aditya Sharma for careful reading of the draft. M. C. gratefully acknowledges Savvas Nesseris and Sachiko Kuroyanagi for the opportunity to undertake a research stay at the Max Planck Institute for Gravitational Physics (Albert Einstein Institute) in Potsdam, for valuable discussions, and for helpful comments on the manuscript. M. C. acknowledges support from the “Ramón Areces” Foundation through the “Programa de Ayudas Fundación Ramón Areces para la realización de Tesis Doctorales en Ciencias de la Vida y de la Materia 2023” and from the research project PID2021-123012NB-C43 and the Spanish Research Agency (Agencia Estatal de Investigación) through the Grant IFT Centro de Excelencia Severo Ochoa No. CEX2020-001007-S, funded by MCIN/AEI/10.13039/501100011033. M. C. also thanks the Max Planck Institute for Gravitational Physics for warm hospitality during the training of this project and the Programa CSIC “iMOVE 2024,” which made this stay possible. S. R. G. is supported by a UKRI Future Leaders Fellowship (Grant No. MR/Y018060/1). This material is based upon work supported by NSF’s LIGO Laboratory which is a major facility fully funded by the National Science Foundation. The computational work for this manuscript was carried out on the computing cluster Saraswati at the Max Planck Institute for Gravitational Physics in Potsdam.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request.

APPENDIX: COMPRESSIBILITY OF LENSED WAVEFORMS

Lensed waveforms are intrinsically more complex than unlensed ones. To quantify the relative complexity of unlensed and microlensed waveforms, we perform a singular value decomposition (SVD) of ensembles of time-aligned frequency-domain signals. Given a data matrix $\mathbf{H}$ whose rows correspond to individual waveforms, the decomposition [93]

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger \quad (\text{A1})$$

defines an orthonormal basis $\{v_k(f)\}$ in waveform space (columns of $\mathbf{V}$). Here $\mathbf{U}$ contains the left singular vectors describing how individual waveforms project onto this basis, while $\mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots)$ contains the singular values ordered by decreasing importance. Truncating the expansion to the first n components yields an approximate reconstruction

$$\tilde{h}^{(n)}(f) = \sum_{k=1}^n c_k v_k(f), \quad (\text{A2})$$

where c_k denote the coefficients. The accuracy of this representation is quantified using the mismatch between $h(f)$ and $\tilde{h}^{(n)}(f)$. As shown in Fig. 9, unlensed waveforms reach mismatch $\mathcal{M} \sim 10^{-4}$ with $n \simeq 200$ basis vectors, whereas microlensed waveforms require approximately twice as many components. Although DINGO does not employ an explicit SVD representation, this diagnostic highlights the increased intrinsic complexity and reduced compressibility of microlensed signals.

One can gain intuition about the reason for the worse scaling of basis vectors for the lensed case with the following argument. We know that not setting the merger

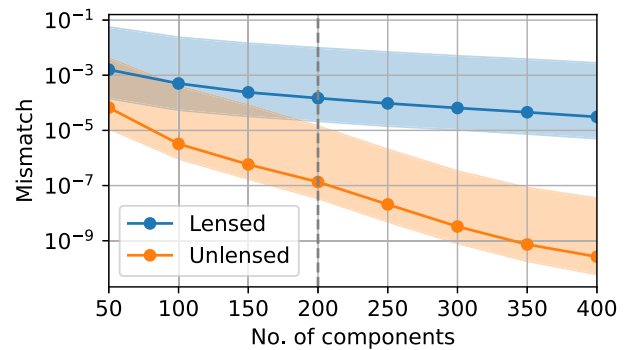


FIG. 9. Mismatch between true waveforms and their reconstruction using a truncated SVD basis. Microlensed waveforms require significantly more components than unlensed ones to reach the same mismatch, reflecting their increased intrinsic complexity. The SVD is used here only as a diagnostic and not in the DINGO analysis.

time to the time of arrival in the detectors (i.e., the merger occurs at $t \neq 0$) drastically increases the number of SVD bases needed to model the waveform, which is one of the reasons for using GNPE [81]. If one thinks of lensing as a superposition of two signals, even though we have time aligned the first signal, the second is free to shift depending on the value of M_{Lz} and y . While the first signal is

standardized, the second is not and the network/SVD has to represent this time shift on its own. For future work, one could consider applying further time shifts to simplify the signal depending values of M_{Lz} and y or increase the complexity of the embedding network that DINGO uses to compress a waveform before passing it to the normalizing flow.

-
- [1] B. P. Abbott *et al.* (Virgo and LIGO Scientific Collaborations), *Phys. Rev. Lett.* **116**, 061102 (2016).
 - [2] B. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **116**, 221101 (2016).
 - [3] M. Isi, M. Giesler, W. M. Farr, M. A. Scheel, and S. A. Teukolsky (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **123**, 111102 (2019).
 - [4] K. Yagi and N. Yunes, *Phys. Rev. D* **95**, 084014 (2017).
 - [5] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), *Phys. Rev. D* **112**, 084080 (2025).
 - [6] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), *Phys. Rev. X* **13**, 011048 (2023).
 - [7] B. A. *et al.* (LIGO Scientific and Virgo Collaborations), *Nature (London)* **551**, 85 (2017).
 - [8] R. Abbott *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), *Astrophys. J.* **949**, 76 (2023).
 - [9] P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational Lenses*, Astronomy and Astrophysics Library (Springer, Berlin, Germany, 1999).
 - [10] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, *Astrophys. J. Lett.* **648**, L109 (2006).
 - [11] N. Aghanim *et al.* (Planck Collaboration), *Astron. Astrophys.* **641**, A8 (2020).
 - [12] S. Vegetti *et al.*, *Space Sci. Rev.* **220**, 58 (2024).
 - [13] M. Bartelmann and P. Schneider, *Phys. Rep.* **340**, 291 (2001).
 - [14] T. Treu, *Annu. Rev. Astron. Astrophys.* **48**, 87 (2010).
 - [15] P. Natarajan, L. L. Williams, M. Bradac, C. Grillo, A. Ghosh, K. Sharon, and J. Wagner, *arXiv:2403.06245*.
 - [16] R. Reyes, R. Mandelbaum, U. Seljak, T. Baldauf, J. E. Gunn, L. Lombriser, and R. E. Smith, *Nature (London)* **464**, 256 (2010).
 - [17] T. E. Collett, L. J. Oldham, R. J. Smith, M. W. Auger, K. B. Westfall, D. Bacon, R. C. Nichol, K. L. Masters, K. Koyama, and R. van den Bosch, *Science* **360**, 1342 (2018).
 - [18] S. Goyal, A. Vijaykumar, J. M. Ezquiaga, and M. Zumalacárregui, *Phys. Rev. D* **108**, 024052 (2023).
 - [19] S. Goyal, K. Haris, A. K. Mehta, and P. Ajith, *Phys. Rev. D* **103**, 024038 (2021).
 - [20] K. C. Wong *et al.* (H0LiCOW Collaboration), *Mon. Not. R. Astron. Soc.* **498**, 1420 (2020).
 - [21] T. Treu, S. H. Suyu, and P. J. Marshall, *Astron. Astrophys. Rev.* **30**, 8 (2022).
 - [22] S. Jana, S. J. Kapadia, T. Venumadhav, and P. Ajith, *Phys. Rev. Lett.* **130**, 261401 (2023).
 - [23] G. P. Smith *et al.*, *Phil. Trans. R. Soc. A* **383**, 20240134 (2025).
 - [24] R. Takahashi and T. Nakamura, *Astrophys. J.* **595**, 1039 (2003).
 - [25] C. Leung, D. Jow, P. Saha, L. Dai, M. Oguri, and L. V. E. Koopmans, *Space Sci. Rev.* **221**, 29 (2025).
 - [26] J. M. Diego, O. A. Hannuksela, P. L. Kelly, T. Broadhurst, K. Kim, T. G. F. Li, G. F. Smoot, and G. Pagano, *Astron. Astrophys.* **627**, A130 (2019).
 - [27] M. H. Y. Cheung, J. Gais, O. A. Hannuksela, and T. G. F. Li, *Mon. Not. R. Astron. Soc.* **503**, 3326 (2021).
 - [28] S. M. C. Yeung, M. H. Y. Cheung, E. Seo, J. A. J. Gais, O. A. Hannuksela, and T. G. F. Li, *Mon. Not. R. Astron. Soc.* **526**, 2230 (2023).
 - [29] A. Mishra, A. K. Meena, A. More, S. Bose, and J. S. Bagla, *Mon. Not. R. Astron. Soc.* **508**, 4869 (2021).
 - [30] M. Oguri and R. Takahashi, *Phys. Rev. D* **106**, 043532 (2022).
 - [31] A. Mishra, A. K. Meena, A. More, and S. Bose, *Mon. Not. R. Astron. Soc.* **531**, 764 (2024).
 - [32] X. Shan, G. Li, X. Chen, W. Zhao, B. Hu, and S. Mao, *Sci. China Phys. Mech. Astron.* **68**, 219512 (2025).
 - [33] R. K. L. Lo, L. Vujeva, J. M. Ezquiaga, and J. C. L. Chan, *Phys. Rev. Lett.* **134**, 151401 (2025).
 - [34] E. Seo, X. Shan, J. Janquart, O. A. Hannuksela, M. A. Hendry, and B. Hu, *Astrophys. J.* **988**, 159 (2025).
 - [35] Z. Su, X. Shan, Z. Lyu, J. Zhang, Y. Liu, S. Mao, and H. Yang, *arXiv:2510.17125*.
 - [36] J. Feldbrugge, U.-L. Pen, and N. Turok, *Ann. Phys. (Amsterdam)* **451**, 169255 (2023).
 - [37] G. Tambalo, M. Zumalacárregui, L. Dai, and M. H.-Y. Cheung, *Phys. Rev. D* **108**, 043527 (2023).
 - [38] X. Shan, G. Li, X. Chen, W. Zheng, and W. Zhao, *Sci. China Phys. Mech. Astron.* **66**, 239511 (2023).
 - [39] S. M. C. Yeung, M. H. Y. Cheung, M. Zumalacárregui, and O. A. Hannuksela, *arXiv:2410.19804*.
 - [40] H. Villarrubia-Rojo, S. Savastano, M. Zumalacárregui, L. Choi, S. Goyal, L. Dai, and G. Tambalo, *Phys. Rev. D* **111**, 103539 (2025).
 - [41] M. Wright and M. Hendry, *Astrophys. J.* **935**, 68 (2022).
 - [42] M. H.-Y. Cheung, K. K. Y. Ng, M. Zumalacárregui, and E. Berti, *Phys. Rev. D* **109**, 124020 (2024).
 - [43] S. Goyal, H. Villarrubia-Rojo, and M. Zumalacárregui, *arXiv:2512.17631*.
 - [44] A. G. Abac *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), *Astrophys. J. Lett.* **993**, L25 (2025).

- [45] J. Janquart *et al.*, *Mon. Not. R. Astron. Soc.* **526**, 3832 (2023).
- [46] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J.* **923**, 14 (2021).
- [47] R. Abbott *et al.* (LIGO Scientific, KAGRA, and Virgo Collaborations), *Astrophys. J.* **970**, 191 (2024).
- [48] A. G. Abac *et al.* (LIGO Scientific, Virgo, and KAGRA Collaborations), [arXiv:2512.16347](https://arxiv.org/abs/2512.16347).
- [49] F. S. Broekgaarden, S. Banagiri, and E. Payne, *Astrophys. J.* **969**, 108 (2024).
- [50] K. K. Y. Ng, K. W. K. Wong, T. Broadhurst, and T. G. F. Li, *Phys. Rev. D* **97**, 023012 (2018).
- [51] S.-S. Li, S. Mao, Y. Zhao, and Y. Lu, *Mon. Not. R. Astron. Soc.* **476**, 2220 (2018).
- [52] G. P. Smith, A. Robertson, G. Mahler, M. Nicholl, D. Ryczanowski, M. Bianconi, K. Sharon, R. Massey, J. Richard, and M. Jauzac, *Mon. Not. R. Astron. Soc.* **520**, 702 (2023).
- [53] A. R. A. C. Wierda, E. Wempe, O. A. Hannuksela, L. é. V. E. Koopmans, and C. Van Den Broeck, *Astrophys. J.* **921**, 154 (2021).
- [54] J. Skilling, *Bayesian Anal.* **1**, 833 (2006).
- [55] J. S. Speagle, *Mon. Not. R. Astron. Soc.* **493**, 3132 (2020).
- [56] G. Ashton *et al.*, *Astrophys. J. Suppl. Ser.* **241**, 27 (2019).
- [57] F. Santoliquido, J. Tissino, U. Dupletsa, M. Branchesi, J. Harms, M. Arca Sedda, M. Dax, A. Kofler, S. R. Green, N. Gupte, I. M. Romero-Shaw, and E. Berti, *Phys. Rev. D* **112**, 103015 (2025).
- [58] Z. Qin, T.-Y. Sun, B.-Y. Li, J.-F. Zhang, X. Guo, and X. Zhang, [arXiv:2505.20996](https://arxiv.org/abs/2505.20996).
- [59] R. Bada-Nerin, O. Bulashenko, O. Gramaxo Freitas, and J. A. Font, *Phys. Rev. D* **111**, 084067 (2025).
- [60] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, *Phys. Rev. Lett.* **127**, 241103 (2021).
- [61] S. R. Green, C. Simpson, and J. Gair, *Phys. Rev. D* **102**, 104057 (2020).
- [62] S. R. Green and J. Gair, *Mach. Learn. Sci. Tech.* **2**, 03LT01 (2021).
- [63] 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).
- [64] N. Gupte *et al.*, *Phys. Rev. D* **112**, 104045 (2025).
- [65] M. Dax, S. R. Green, J. Gair, N. Gupte, M. Pürrer, V. Raymond, J. Wildberger, J. H. Macke, A. Buonanno, and B. Schölkopf, *Nature (London)* **639**, 49 (2025).
- [66] Q. Hu, [arXiv:2507.05209](https://arxiv.org/abs/2507.05209).
- [67] J. Urrutia and V. Vaskonen, *Mon. Not. R. Astron. Soc.* **509**, 1358 (2021).
- [68] G. Tambalo, M. Zumalacárregui, L. Dai, and M. H.-Y. Cheung, *Phys. Rev. D* **108**, 103529 (2023).
- [69] S. Savastano, G. Tambalo, H. Villarrubia-Rojo, and M. Zumalacárregui, *Phys. Rev. D* **108**, 103532 (2023).
- [70] H. Gil Choi, S. Jung, P. Lu, and V. Takhistov, *Phys. Rev. Lett.* **133**, 101002 (2024).
- [71] M. Zumalacárregui, [arXiv:2404.17405](https://arxiv.org/abs/2404.17405).
- [72] L. Vujeva, J. M. Ezquiaga, D. Gilman, S. Goyal, and M. Zumalacárregui, [arXiv:2510.14953](https://arxiv.org/abs/2510.14953).
- [73] G. Brando, S. Goyal, S. Savastano, H. Villarrubia-Rojo, and M. Zumalacárregui, *Phys. Rev. D* **111**, 024068 (2025).
- [74] C. R. Keeton, [arXiv:astro-ph/0102341](https://arxiv.org/abs/astro-ph/0102341).
- [75] D. L. Jow, S. Foreman, U.-L. Pen, and W. Zhu, *Mon. Not. R. Astron. Soc.* **497**, 4956 (2020).
- [76] G. Hinshaw and L. M. Krauss, *Astrophys. J.* **320**, 468 (1987).
- [77] J. F. Navarro, C. S. Frenk, and S. D. M. White, *Astrophys. J.* **490**, 493 (1997).
- [78] E. Thrane and C. Talbot, *Pub. Astron. Soc. Aust.* **36**, e010 (2019).
- [79] K. Akiyama *et al.* (Event Horizon Telescope Collaboration), *Astrophys. J. Lett.* **875**, L1 (2019).
- [80] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, [arXiv:1912.02762](https://arxiv.org/abs/1912.02762).
- [81] M. Dax, S. R. Green, J. Gair, M. Deistler, B. Schölkopf, and J. H. Macke, [arXiv:2111.13139](https://arxiv.org/abs/2111.13139).
- [82] G. Pratten *et al.*, *Phys. Rev. D* **103**, 104056 (2021).
- [83] A. e. a. Buikema, *Phys. Rev. D* **102**, 062003 (2020).
- [84] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/lvwm-d7fm> for an unlensed injection analyzed with unlensed DINGO.
- [85] J. C. L. Chan, J. M. Ezquiaga, R. K. L. Lo, J. Bowman, L. M. Zertuche, and L. Vujeva, [arXiv:2512.16916](https://arxiv.org/abs/2512.16916).
- [86] A. Kofler, M. Dax, S. R. Green, J. Wildberger, N. Gupte, J. H. Macke, J. Gair, A. Buonanno, and B. Schölkopf, [arXiv:2512.02968](https://arxiv.org/abs/2512.02968).
- [87] S. Goyal, H. Villarrubia-Rojo, and M. Zumalacárregui, [arXiv:2512.17631](https://arxiv.org/abs/2512.17631).
- [88] X. Shan, H. Yang, and S. Mao, [arXiv:2512.19118](https://arxiv.org/abs/2512.19118).
- [89] Q. Hu, [arXiv:2507.05209](https://arxiv.org/abs/2507.05209).
- [90] N. Rao, A. Mishra, A. Ganguly, and A. More, *Phys. Rev. D* **113**, 023054 (2026).
- [91] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith, *Nat. Phys.* **18**, 112 (2021).
- [92] J. C. Chan, L. M. Zertuche, J. M. Ezquiaga, R. K. Lo, L. Vujeva, and J. Bowman, *Phys. Rev. D* **113**, 024041 (2026).
- [93] P. C. Hansen, *BIT* **27**, 534 (1987).