

# Learning Stellar Fields with Microlensed Gravitational-Wave Events

Zhaocqi Su,<sup>1,2</sup> Xikai Shan , Zhenwei Lyu,<sup>4</sup> Shude Mao,<sup>5</sup> and Huan Yang <sup>3,\*</sup>

<sup>1</sup>*College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China*

<sup>2</sup>*Department of Automation, Tsinghua University, Beijing 100084, China*

<sup>3</sup>*Department of Astronomy, Tsinghua University, Beijing 100084, China*

<sup>4</sup>*Leicester International Institute, Dalian University of Technology, Panjin 124221, China*

<sup>5</sup>*Department of Astronomy, Westlake University, Hangzhou 310030, Zhejiang Province, China*

(Dated: September 2024)

Strongly lensed gravitational waves may pass through the star field of the lensing galaxy with additional modulation (on both phase and amplitude) due to gravitational fields of stars near the line of sight, i.e. the microlensing effect. These microlensed waveforms may depend on the mass and location of tens or more most relevant stars, so that their deterministic reconstruction from data is computationally prohibitive. We classify the detection and parameter estimation of such events as *non-deterministic inference problem* and propose a solution with the implementation of normalizing flows. As a first step, we show that 14% of microlensed events can be detected with significance  $\geq 2\sigma$  in the 3G era, with the chosen microlensing parameters correlated with the density of the underlying star field. This approach opens the door of probing microlensing effects and the properties of the underlying star fields, with similar construction may be applied to other nondeterministic inference problems, such as detecting post-merger gravitational waves from binary neutron star coalescence and signals from core-collapse supernovae.

## INTRODUCTION

Strongly-lensed gravitational waves are emerging as novel tracers with unique advantages in cosmological and astrophysical studies. For instance, millisecond-level time-delay measurements from strongly-lensed binary black holes (SLBBHs) offer the potential for a more precise estimation of the Hubble parameter. Additionally, SLBBHs can improve the localization precision of binary black holes mergers [1, 2], and serve as a novel platform for testing general relativity [3–5].

Moreover, gravitational wave lensing is not always adequately described by the geometric optics approximation. When the wavelength of the gravitational waves is comparable to the Schwarzschild radius of the lens, wave-optics effects become significant. In this scenario, a characteristic oscillatory behavior, driven by diffraction, manifests itself in the gravitational wave waveform as the frequency sweeps upward. This phenomenon is common and inevitable for lensed images that pass through a field of stellar-mass microlenses within the strong lensing galaxy [6–9]. If these oscillations can be faithfully detected and accurately characterized, the mass distribution of the microlenses, ranging from intermediate-mass black holes to substellar compact objects, may be uncovered to a certain extent. This provides an unprecedented opportunity for studying faint or isolated compact object populations in galaxies. In this study, the first of its kind, we propose an artificial intelligence-based method to identify microlensed binary black holes (BBHs) via their distinctive lensing signatures, and to subsequently infer the parameters associated with the stellar field.

The microlensing waveform encodes the high-dimensional structure of the stellar field, but extracting

all this information is both computationally infeasible and scientifically unnecessary. Instead, it suffices to measure a few characteristic observables of the waveform, which in turn can be used to infer the properties of the stellar field. The statistical significance of these characteristic observables also allows a quantitative measure for the detection statistics of the microlensing effects. However, notice that the same set of characteristic observables can arise from many different stellar-field realizations, meaning that a waveform containing only the characteristic observables is *intrinsically nondeterministic*.

Detection and parameter estimation with nondeterministic waveforms differ fundamentally from deterministic waveforms, such as those employed in studies of probing binary black holes and neutron stars in Advanced LIGO-Virgo events [10]. Despite these differences, the inference problem with non-deterministic waveforms is expected to have broad applications in gravitational wave astronomy. Besides the microlensing measurement problem, intrinsic stochasticity may also affect waveforms from post-merger neutron stars and core-collapse supernovae, due to processes such as MHD turbulence, uncertain equations of state, or incomplete theoretical modeling [11]. Therefore, it is both observationally and theoretically important to identify pathways towards solving the nondeterministic inference problem.

Focusing on the important scenario of stellar-field microlensing, we take an initial step in this direction with neural-network techniques. We argue that the normalizing-flow framework naturally yields posterior distributions for selected characteristic observables of a nondeterministic waveform. To validate this approach, we test this proposal with a simple example using the binary black hole waveform and treating the chirp mass

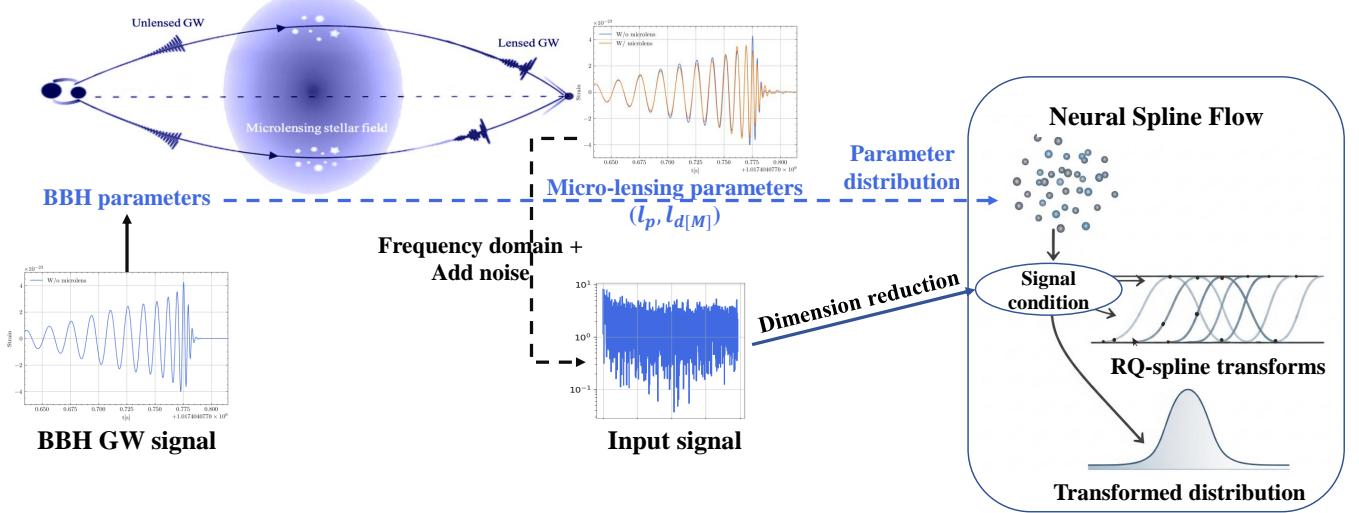


FIG. 1. Microlensed BBHs and the detection method with neural networks.

as a latent variable. In this example, where Monte Carlo-based parameter inference remains tractable, the nondeterministic inference using normalizing flow produces consistent results as the Monte-Carlo method. This finding motivates us to further apply the same technique for more complex microlensed waveforms.

We prepare  $10^8$   $10^6$ ? microlensed waveforms by statistically sampling an equal number of stellar-field realizations, all drawn from the same stellar mass function (see Fig. 1). The computation of each waveform, based on the diffraction integral, is significantly accelerated using the method described in [], with runtime ranging from ??seconds to ?? minutes depending on the complexity of the stellar configurations. For a generic microlensed waveform, we define four characteristic variables, chosen to reflect the magnification strength of microlensed images and the frequency-dependent deviations from unlensed waveforms. These waveforms are randomly assigned to binary black hole waveforms in vacuum, which are subsequently used to train a neural network based on normalizing flows. Notice that the micro-“lensing” labels of each waveform only contain the four chosen lensing variables, so that the intrinsic randomness of the waveforms is encoded in the training process.

Training converges after  $\sim 300$  epochs, where the neural network is able to generate posterior distributions for the selected lensing parameters. Assuming 3G detector sensitivities and realistic astrophysical distributions, we observe that approximately 14% (8%) of strongly lensed events will show deviation  $2\sigma$  ( $3\sigma$ ) with waveforms without microlensing, which roughly corresponds to  $\mathcal{O}(10)$  per year. These “golden events” can be further exploited to probe properties of the intervening stellar field, such as microlens mass distribution and strong-lensing mag-

nification. Although more room for exploration and optimization is expected, this framework enables the possibility of probing microlensed BBHs through stellar fields, with potential extension to other non-deterministic problems in the future.

## RESULTS

### Non-deterministic waveform

GWs passing through a stellar field are likely to be micro-lensed by many nearby stars, such that the resulting waveform can be expressed as  $h(\theta_{\text{GR}}, \theta_*)$ , where  $\theta_{\text{GR}}$  denotes the binary parameters and  $\theta_*$  represents the mass and positions of all relevant stars. However, since the dimension of  $\theta_*$  (number of stars in the simulation box) is likely  $\geq \mathcal{O}(10^4)$ , it is computationally prohibitive to perform parameter estimation (PE) on  $\theta_*$ . A more practical approach is to identify a subset of lensing parameters,  $\theta_s \in \theta_*$ , and perform joint PE with  $\theta_{\text{GR}}$ . This process can be viewed as inference with nondeterministic waveforms, as  $(\theta_{\text{GR}}, \theta_s)$  alone do not uniquely determine the microlensed waveform. The corresponding posterior distribution should be given by

$$P(\theta_{\text{GR}}, \theta_s) = \int d\theta_{*-s} P(\theta_{\text{GR}}, \theta_*) \quad (1)$$

where  $\theta_{*-s}$  denotes the remaining lensing parameters such that  $(\theta_{*-s}, \theta_s) = \theta_*$ , although  $P(\theta_{\text{GR}}, \theta_*)$  is difficult to obtain with Monte-Carlo methods.

We propose applying the normalizing flow framework to directly compute the joint distribution  $P(\theta_{\text{GR}}, \theta_s)$ , as a simulation-based inference problem. Firstly, during

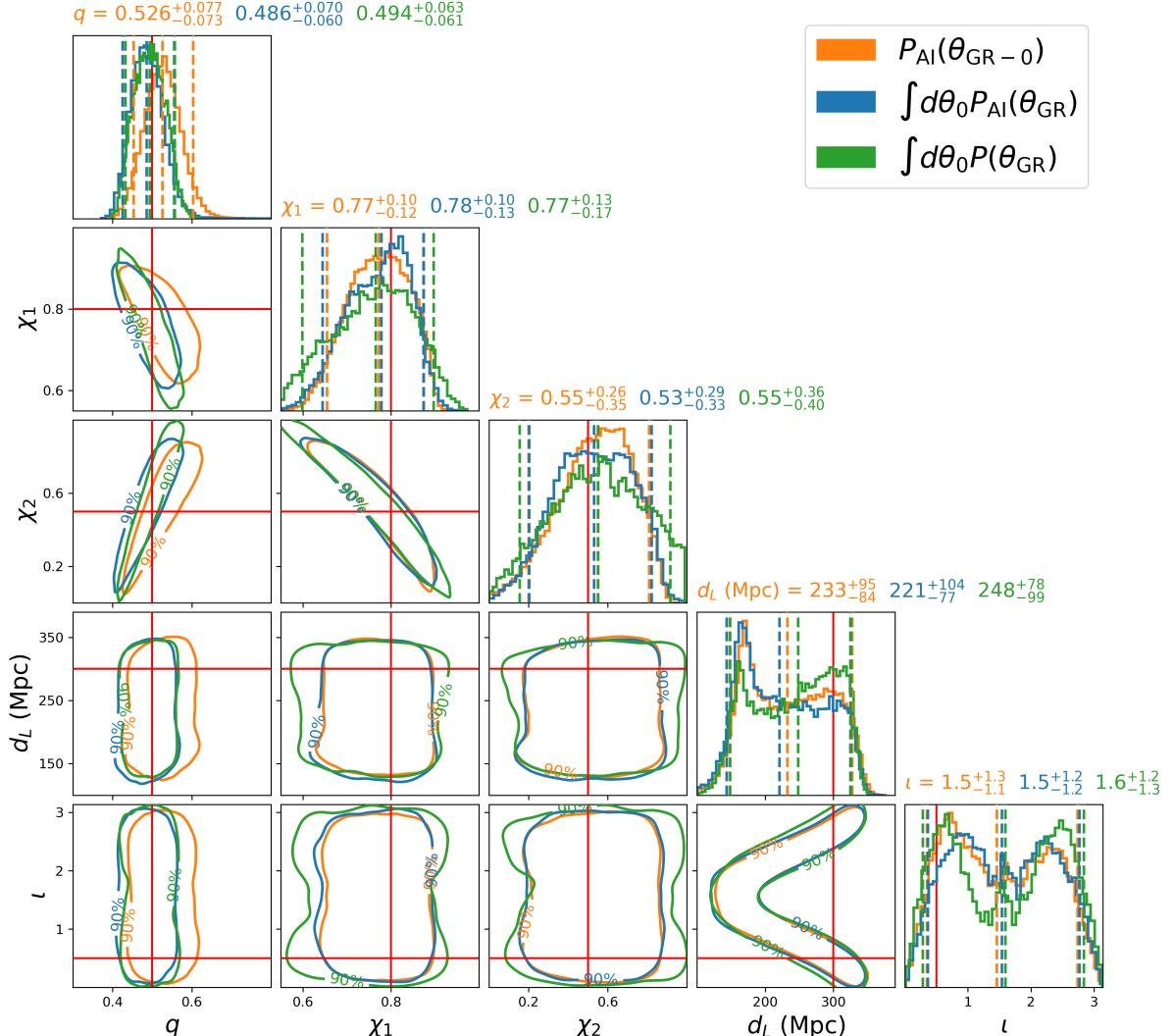


FIG. 2. The posterior distributions of the mass ratio  $q$ , aligned spins of the primary and secondary BHs ( $\chi_1$  and  $\chi_2$ , respectively), the luminosity distance  $d_L$ , and inclination angle  $i$  are shown. The orange curve labeled " $P_{\text{AI}}(\theta_{\text{GR}-0})$ " shows the Neural Posterior Estimation (NPE) result without chirp mass as an inference parameter. The blue curve labeled " $\int d\theta_0 P_{\text{AI}}(\theta_{\text{GR}})$ " represents the NPE result with chirp mass included as an inference parameter. The green curve labeled " $\int d\theta_0 P(\theta_{\text{GR}})$ " corresponds to the result obtained from a traditional MCMC simulation using the `emcee_pt` sampler embedded in the PyCBC package [10]. The solid curves in the 2D contour plots represent the 90% credible regions, while the vertical dashed lines indicate the two-sided 90% confidence intervals. The solid red lines mark the injected (true) parameter values.

training, the detector noise can be randomly simulated using the real detector noise added to each training waveform, i.e.  $d = n + h$  [11]. This enables the normalizing flow to handle inference problems involving stochastic detector noise. Since the neural network is agnostic to the physical origin of the noise, only recognizing its presence in the data, it should likewise be able to deal with sce-

narios where randomness originates from the waveform itself. Secondly, the normalizing flow is also capable of marginalizing over unconsidered parameters during inference. This is because the normalizing flow learns the full joint distribution of all parameters during training. At the inference stage, marginalization over any subset of parameters can be performed efficiently via integration

over the learned distribution, without requiring retraining or explicit sampling over the marginalized parameters.

As a proof of principle, we consider the BBH waveform with GR parameter  $\theta_{\text{GR}}$  and split it into  $(\theta_0, \theta_{\text{GR}-0})$ . Here,  $\theta_0$  (chosen as the chirp mass here) is treated as  $\theta_{*-s}$ , which is not explicitly labeled in neural network training. The normalizing flow then yields the posterior distribution of  $P_{\text{AI}}(\theta_{\text{GR}-0})$ , which can be validated by checking whether it satisfies

$$P_{\text{AI}}(\theta_{\text{GR}-0}) \approx P(\theta_{\text{GR}-0}) = \int d\theta_0 P(\theta_{\text{GR}}). \quad (2)$$

Notice that in this example  $P(\theta_{\text{GR}})$  is known through Monte-Carlo methods, so comparison can be explicitly performed.

Indeed, as shown in Fig. 2, the posterior distribution  $P_{\text{AI}}(\theta_{\text{GR}-0})$  is largely consistent with both  $\int d\theta_0 P(\theta_{\text{GR}})$  and  $\int d\theta_0 P_{\text{AI}}(\theta_{\text{GR}})$ , where the training of  $P_{\text{AI}}(\theta_{\text{GR}})$  corresponds to deterministic waveforms. Although this promising consistency has only been verified in a specific scenario, it suggests that a similar relation may still hold for more general settings, which motivates the application of normalizing flow to infer the posterior distribution of lensing parameters.

## Data and training

Lensing data are generated by multiplying the source gravitational waveform from BBHs with the microlensing magnification waveform in the frequency domain. Unlike DINGO [11], where the gravitational waveform is specified deterministically by the BBH parameters, these microlensing gravitational wave (MLGW Mock) data introduce additional stochasticity. The microlensing signal, arising from a random distribution of stars, cannot be uniquely characterized by the few chosen lensing parameters alone, resulting in high-dimensional variability. Consequently, the forward model is intrinsically probabilistic: identical BBH and lensing parameters can produce distinct waveforms due to micro-lensing effects, significantly increasing the modeling complexity.

For the dataset, the binary black hole (BBH) signals are generated using LALSimulation [12], while the microlensing signal is produced through a dedicated pipeline, detailed in the “Waveform preparation” section of the Appendix. To maintain physical consistency, the redshift of the lens is constrained to be smaller than that of the BBH source. This is achieved through acceptance-rejection sampling applied during the training and testing processes. Other processing steps, including whitening and noise addition, follow the same procedures as in DINGO [11]. For method validation, we employ two Cosmic Explorer (CE) detectors and remove the glitch in the noise power spectrum density.

For the micro-lensing signal, we use 4 different parameters for parametrization:  $l_p$ ,  $l_{d10}$ ,  $l_{d60}$  and  $l_{d120}$ , as shown in Fig. 3. The parameter  $l_p$  is calculated from the sum of the 10 highest peaks of the time-domain magnification factor,  $\tilde{F}(t, \mathbf{y})$ , defined as:

$$\tilde{F}(t, \mathbf{y}) = \int_{-\infty}^{\infty} dt' \delta[t - t'] = \frac{|dS|}{dt}. \quad (3)$$

In the context of microlensing, each peak in the time-domain magnification corresponds to a single microlensing image. The physical meaning of a peak’s height can be understood by analogy to the study of the microlensing effect within strong lensing images. In such studies, the strong lensing potential is locally expanded around the strong lensing images using the convergence  $\kappa$  and shear  $\gamma$  to incorporate strong lensing contributions into the microlensing stellar field.

Similarly, in our study, we can expand the total potential from both microlensing and strong lensing around the microlensing images, which correspond to the peaks in the time-domain magnification. This approach allows us to analogize the time-domain magnification of an individual microlensing image to that of an individual strong lensing image, which is given by  $\tilde{F}(t)_i = -2\frac{\sqrt{\mu_i}}{k} \ln(t - t_i) + C$  [6]. Here,  $k = 4GM_{\text{micro}}(1 + z_L)/c^3$ ,  $\mu_i$  is the magnification of the  $i$ -th microlensing image, and  $C$  is a constant from the numerical truncation that does not affect the frequency-domain magnification factor. Therefore, the 10 highest peaks selected from  $\tilde{F}(t, \mathbf{y})$  represent the 10 most magnified microlensing images.

The  $l_{d[M]}$  is defined as:

$$l_{d[M]} = \Im \left[ \frac{\langle h_1 | h_1 - h_2 \rangle}{\sqrt{\langle h_1 | h_1 \rangle \langle h_1 - h_2 | h_1 - h_2 \rangle}} \right], \quad (4)$$

where  $\Im$  denotes the imaginary part of the complex quantity. The term  $\langle \cdot | \cdot \rangle$  represents the noise-weighted inner product, defined as:

$$\langle h_1 | h_2 \rangle = 4 \int_{f_{\text{low}}}^{f_{\text{high}}} df \frac{h_1^*(f) h_2(f)}{S_n(f)}, \quad (5)$$

where  $S_n(f)$  is the single-sided power spectral density of the detector noise, and  $*$  denotes the complex conjugate. In this context, signal 1 is the unlensed waveform ( $h_U$ ), and signal 2 is the macro and microlensed waveform ( $h_L$ ).

We introduce this new definition to replace the traditional mismatch parameter for two main reasons. First, the conventional mismatch parameter is always greater than zero, giving it a single-sided probability distribution. Training a network on such a “biased” distribution may lead to biased parameter estimation. In contrast,  $l_{d[M]}$  follows a distribution centered at zero, where zero represents the null case (i.e., no microlensing signature), as shown in Supplementary Figure 6. This property leads to better performance in both parameter estimation and

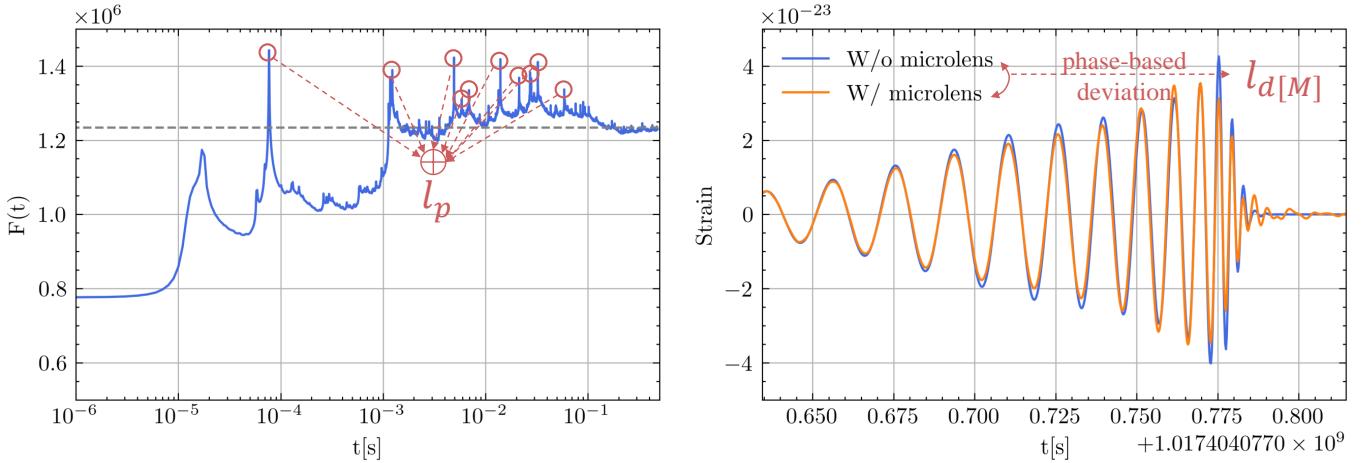


FIG. 3. Micro-lensing parameters for network inference.

microlensing event selection. Second, this new parameter effectively quantifies the phase-based deviation induced by the microlensing effect. A stronger microlensing effect leads to a greater difference between the lensed and unlensed waveforms ( $h_1 - h_2$ ). This, in turn, increases the value of  $\Im[\langle h_1 | h_1 - h_2 \rangle]$ , resulting in a larger value for  $l_{d[M]}$ .

In this work, we employ three distortion parameters,  $l_{d10}$ ,  $l_{d60}$ , and  $l_{d120}$ , which correspond to distortions in binary black hole systems with component masses of  $10 M_\odot - 10 M_\odot$ ,  $60 M_\odot - 60 M_\odot$ , and  $120 M_\odot - 120 M_\odot$ , respectively. These parameters enable a more comprehensive description of the distortions in low-mass, intermediate-mass, and high-mass binary black hole events.

### PE results and implications

Figure 4 (left, orange) presents a representative parameter estimation (PE) result for an injection containing a distinct microlensing signal. The figure displays the marginalized one- and two-dimensional posterior distributions for both the source gravitational-wave and lensing parameters, with 50% and 90% credible regions. Here, we focus on the PE results for the lensing parameters; comprehensive PE results are provided in the Appendix. For the lensing parameters, deviations of the posterior distributions from zero indicate the statistical significance of microlensing. Both the magnitude of the shift from zero and the width of the posterior contribute to this interpretation. To quantitatively assess the calibration and reliability of the inferred posteriors across the population, we refer to the p-p plot in the Appendix, which provides a consistency check between injected and recovered parameters over multiple realizations.

In generic microlensing (ML) events, strong lensing

(SL) components also exist. However, our primary interest lies in detecting the distinct signatures introduced by the microlensing effect. To isolate these effects, we design a null test by preparing comparison events that share the same strong-lensing component but without microlensing. Specifically, we compare the parameter posteriors obtained from two injections: (1) a waveform that includes both microlensing and strong-lensing effects, as described previously; and (2) the same BBH waveform without microlensing, scaled by a strong-lensing magnification factor  $\sqrt{\frac{1}{1-2\kappa}}$ . This ensures that both injections share the same SL amplitude modification, isolating the ML-induced waveform fluctuations. Figure 4 (left) compares posterior distributions inferred from two injections: one with microlensing (orange) and one with only strong lensing (blue). For the strong-lensing-only case, where the waveform is scaled by a macro-lens magnification factor without microstructure, the posterior distributions of the lensing parameters (e.g.,  $l_{d10}$ ,  $l_{d60}$ , and  $l_{d120}$ ) are peaked around zero with a comparatively narrower width, consistent with an absence of microlensing signatures in the data. In contrast, the microlensed injection (orange) exhibits posterior distributions that are consistently shifted away from zero across all lensing parameters. The one-dimensional posteriors show distinct peaks that are either non-overlapping with ( $l_p$ ) or partially overlapping in the tails of ( $l_{d[M]}$ ) those from the SL-only case (blue), indicating that the network has captured additional waveform features beyond amplitude rescaling. For the two-dimensional posterior distributions, in panels involving the lensing peak parameter  $l_p$ , such as  $(l_p, l_{d10})$ ,  $(l_p, l_{d60})$ , and  $(l_p, l_{d120})$ , the 90% credible contours of the microlensed (orange) and SL-only (blue) cases are clearly separated, while the posteriors involving only deviation parameters, such as  $(l_{d10}, l_{d60})$  and  $(l_{d60}, l_{d120})$ , partially overlap but still exhibit visibly

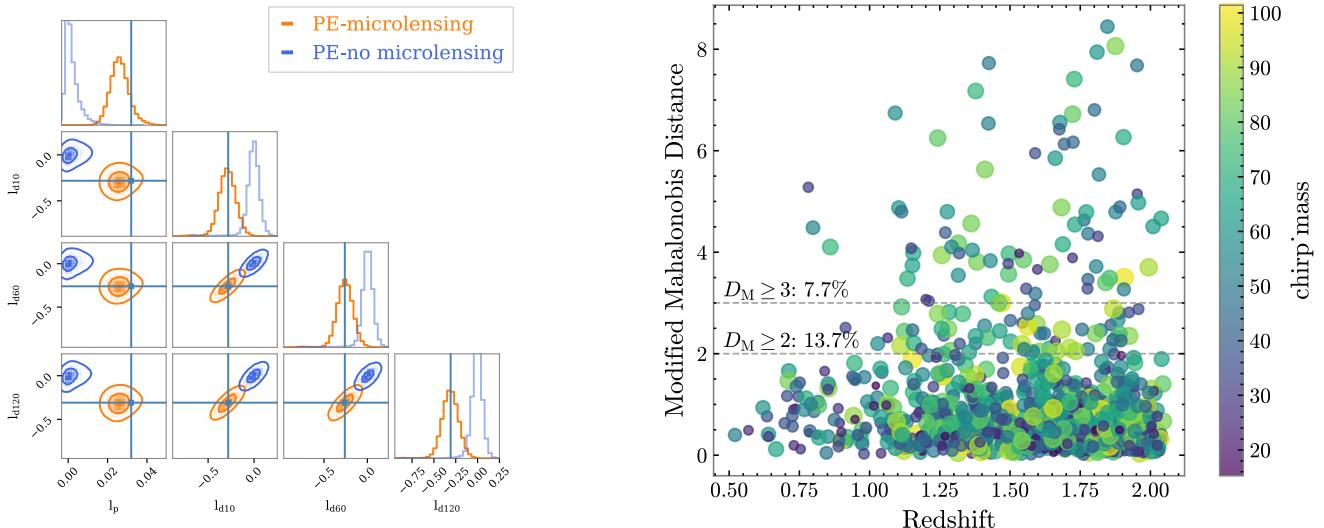


FIG. 4. Detection of microlensing signal. Left: a sampled PE result from microlensing and zero-microlensing injection. Right: the modified Mahalanobis distance distribution regarding to the redshift and the chirp mass of the BBH.

different orientations and centers. This indicates that in this example,  $l_p$  provides the most distinctive signal of microlensing, while the deviation parameters contribute additional but subtler discriminative power through their correlated structure.

Nonetheless, not all microlensing waveforms are easily distinguishable, as typical microlensing signatures are relatively weak compared to the BBH signal. It is therefore essential to assess the distinguishability of microlensed signals quantitatively. To this end, we introduce a modified Mahalanobis distance to quantify the separation between two posterior distributions over the lensing parameters. While the standard Mahalanobis distance measures the distance between a single point and a multivariate distribution, we adapt it here to evaluate the distance between two distributions, defined as:

$$D_M = \sqrt{(\mu_1 - \mu_2)^\top (\Sigma_1 + \Sigma_2)^{-1} (\mu_1 - \mu_2)}, \quad (6)$$

where:

- $\mu_1, \mu_2$  are the mean vectors of lensing parameters in the two posterior distributions,
- $\Sigma_1, \Sigma_2$  are the corresponding covariance matrices of lensing parameters.

This modified Mahalanobis distance measures the separation between the two posterior means, normalized by their combined uncertainty. It serves as a scalar indicator of the distinguishability between two distributions, and higher values suggest the microlensing signal has induced statistically significant shifts in the inferred parameters. Geometrically, it corresponds to the distance between the distribution centers measured in the natural metric defined by the total covariance  $(\Sigma_1 + \Sigma_2)$ . A

value of  $D_M = k$  indicates that the two posterior means are separated by  $k$  standard deviations in the joint uncertainty space. This acts as a statistical discriminability measure between the two cases. For approximately Gaussian posteriors, values above 2 generally correspond to distinguishable waveform features beyond the noise level.

To evaluate the performance of our model under astrophysical distribution [point to SM](#), we sample 1,000 instances and calculate lensing  $D_{M\text{-modified}}$  for each sample. Fig. 4 right shows the  $D_{M\text{-modified}}$  distributions of 1000 samples, with the x axis indicating the redshift of the source BBH, while the size and color bar represent the chirp mass. 7.7% of the samples have  $D_{M\text{-modified}} > 3$  and 13.7% of the samples have  $D_{M\text{-modified}} > 2$ . Additionally, one may argue that microlensing is only detectable for specific event types, such as low-redshift (high-SNR) or low-chirp-mass events, where microlensing effects are more pronounced in the high-frequency band. However, the right panel of Figure 4, which plots the Mahalanobis distance against the redshift and the chirp mass, shows no strong preference for either of the two.

There is a possible explanation as follows. Although low redshift events have a statistically higher SNR, the microlensing effect is modulated by a factor of  $(1 + z_l)$ , as seen in Eq. (9). This means that a microlens at a higher redshift effectively behaves like a more massive one, enhancing its signature. Consequently, our detection efficiency is not strongly biased toward low-redshift events. A similar trade-off exists for the chirp mass: lighter events exhibit more significant microlensing effects but have a lower SNR.

Although the current model does not explicitly predict physical lensing parameters such as the convergence

$\kappa$ , which reflects the total surface mass density in the lens galaxy, and  $\kappa_s$ , which isolates the portion attributable to compact objects such as stars, and is therefore more directly responsible for generating microlensing signatures, we investigate their correlation with our proposed microlensing discriminators. In particular, we examine how the modified Mahalanobis distance, which serves as our primary metric for quantifying microlensing detectability, varies with  $\kappa$  (x-axis) and  $\kappa_s$  (y-axis), as shown in Fig. 5 left. The results indicate a clear trend: events with higher values of  $\kappa$  and, more prominently,  $\kappa_s$  tend to exhibit larger Mahalanobis distances. This implies that microlensing signals become more distinguishable as the contribution from the stellar population increases, even though these convergence parameters were not explicitly included in the model or training data. This provides evidence that the microlensing-sensitive latent representation learned by our network carries predictive power for physical quantities like  $\kappa_s$ , despite not being directly trained on them. The correlation suggests that the learned lensing parameters encode information that is physically related to the underlying stellar field.

To further explore this relation, we also visualize the inferred peak lensing parameter  $l_p$  across the same  $\kappa$ - $\kappa_s$  plane (Fig. 5, right). A similar increasing trend is observed:  $l_p$  tends to rise systematically with increasing  $\kappa$  and  $\kappa_s$ , suggesting that the network's learned representation of lensing structure reflects physical features of the star field. Together, these results demonstrate that our model not only captures microlensing effects at the waveform level, but also infers higher-level physical trends associated with the lens environment.

In contrast, other lensing parameters such as  $l_{d[M]}$  do not show strong monotonic correlation with  $\kappa$  or  $\kappa_s$ . However, this does not diminish their relevance. While  $\kappa$  and  $\kappa_s$  describe the lensing surface mass density and the stellar contribution, other observables like  $l_{d[M]}$  may encode complementary dimensions of microlensing, such as phase-based fluctuations or fine-scale patterns associated with the stellar mass distribution. This highlights the multidimensional nature of microlensing and points toward the benefit of incorporating additional physical descriptors in future modeling. Future studies can further develop this connection by refining the selection of lensing features and introducing more physically informed parameters to describe the mass distribution. Although this result remains preliminary, it demonstrates consistency with gravitational lensing theory and motivates deeper exploration of how microlensing observables can reveal the properties of stellar populations in lensing galaxies.

## DISCUSSION

The successful identification of strongly lensed gravitational waves and their inherent microlensing signatures

is a crucial prerequisite for their application in cosmological and astrophysical studies. A significant challenge and limiting factor is the non-deterministic nature of the lensed waveform for a given strong lensing convergence  $\kappa$ , shear  $\gamma$ , microlensing field density  $\kappa_s$ , and microlens mass function. In this paper, we address this challenge employing a normalizing flow, marking the first application of this machine learning technique to the problem.

Our result implies that these surrogate parameters can effectively differentiate the microlensing effect from a null background. To quantify this distinguishability, we calculate the modified Mahalanobis distance between the posterior contours for cases with and without microlensing. The result in right panel of Figure 4 shows that 7.7% and 13.7% of SLGW events exhibit distinguishable microlensing signatures at a  $3\sigma$  and  $2\sigma$  confidence level, respectively. This means that approximately 3 to 77 events per year will have a  $3\sigma$  detection in the 3G detector era. This number depends on the strong lensing galaxy population and the BBH population model, which predict approximately 40 to 1000 SLGW per year [13]. In addition, although type-II strongly-lensed events are not included in this analysis, they generally exhibit stronger and more easily identifiable microlensing effects, and likely larger fraction of detectable events.

Note that the surrogate parameters are designed to better describe the microlensing signatures in the waveform. They are not commonly used “lensing variables”, such as  $\kappa, \kappa_s$ , but they are correlated with them. This correlation is evident in Fig. 5, which shows that the bin-averaged peak value are positively correlated with both the strong lens convergence ( $\kappa$ ) and in particular, the microlensing density ( $\kappa_s$ ) which is also a measure of the density of the stellar field. This finding is physically intuitive, as the statistical significance of the microlensing effect is expected to increase with higher values of  $\kappa$  and  $\kappa_s$ . This correlation also suggests the possibility of constraining these physical parameters through a hierarchical Bayesian inference framework, or a dedicated mapping between waveform surrogate variables and more detailed mass distribution in the microlensing field. There are likely to be more effective sets of parameters to separately characterize the microlensing waveform modulation and the lensing field. Constructing an informative mapping between them will be an interesting direction to pursue for future studies.

These findings also crucially depend on the assumed detector sensitivity. By scaling our results from two to three third-generation (3G) detectors, we project that the fraction of distinguishable microlensing events should increase to  $\sim 10\%$  ( $3\sigma$ ) and  $\sim 17\%$  ( $2\sigma$ ). This neural network-based method shows an improvement over the mismatch test technique proposed in [14], which could identify 10% events at a confidence level 90% ( $< 2\sigma$ ). More importantly, the normalizing flow allows parameter estimation of the lensing waveform, which is the critical

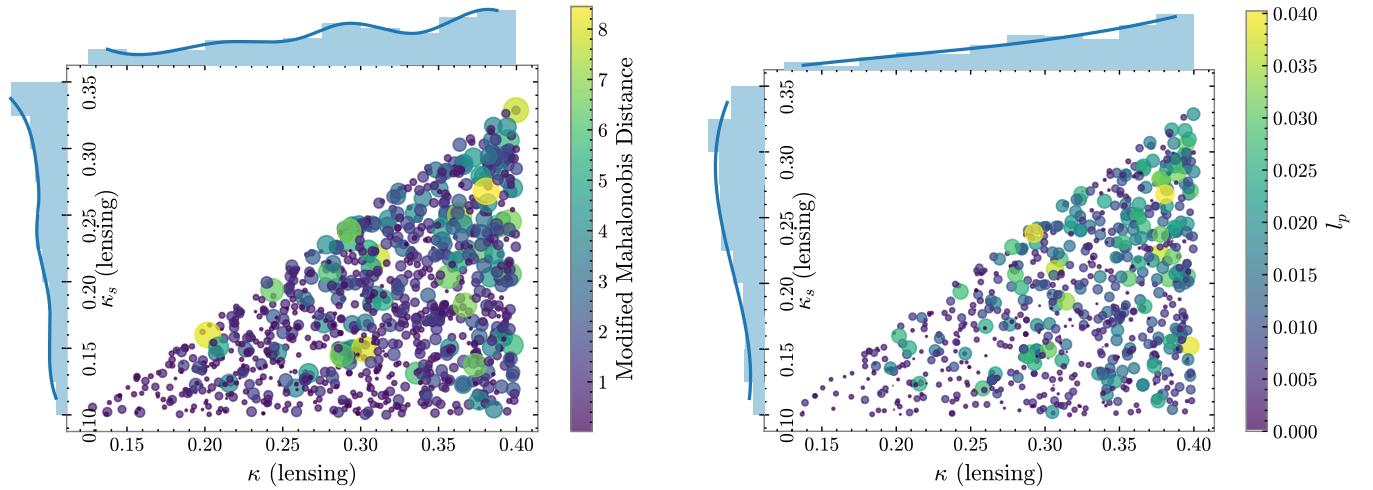


FIG. 5. Modified Mahalanobis distance distribution and  $l_p$  distribution with  $\kappa$  and  $\kappa_s$

first step toward measuring the lensing field.

This successful implementation of the normalizing flow method sheds new light on more general non-deterministic waveform inference problems in gravitational wave astronomy. There are important astrophysical systems, such as post-merger neutron stars, core-collapse supernovae, and extreme mass-ratio inspirals in turbulent Active Galactic Nuclei (“Wet EMRIs”, see [15, 16]), which all exhibit nondeterministic waveforms from various origins. The traditional Monte-Carlo method is unsuitable for probing these systems, but the normalizing flow approach has shown its unique advantages for these simulation-based inference problems. There are still major difficulties before a straightforward implementation on all these problems, such as generating sufficient number of training waveforms, but the above analysis already manifests itself as one of the most promising avenues to explore.

## ACKNOWLEDGMENTS

We thank

## METHOD

### Data Preparation

#### *Training Data Set*

Our training dataset was generated by uniformly sampling the following parameters: strong lensing convergence ( $\kappa$ ), stellar microlensing convergence ( $\kappa_s$ ), source redshift ( $z_s$ ), and lens redshift ( $z_l$ ).

The parameter ranges were defined as follows:

- $\kappa \propto U(0.1, 0.4)$
- $\kappa_s \propto U(0.1, \kappa/1.2)$
- $z_s \propto U(0.1, 2.0)$
- $z_l \propto U(0.15, z_s)$

The strong lensing convergence, denoted by  $\kappa$ , describes the dimensionless surface mass density at the location of a strong lensing image. It is defined as the ratio of the surface mass density,  $\Sigma$ , to the critical surface density,  $\Sigma_{\text{crit}}$ :

$$\kappa = \frac{\Sigma}{\Sigma_{\text{crit}}}. \quad (7)$$

The convergence of the microlensing stellar field, denoted by  $\kappa_s$ , is defined as:

$$\kappa_s = \frac{N\pi}{S}, \quad (8)$$

where  $N$  is the number of microlenses and  $S$  is the area of the region, normalized by the square of the Einstein radius corresponding to the average mass of the microlenses.

The upper limit for  $\kappa_s$  ensures that the total convergence from discrete objects ( $1.2\kappa_s$ , assuming a 20% remnant fraction) does not exceed the total convergence  $\kappa$ .

Additionally, we assume that the strong lensing shear equals the convergence ( $\gamma = \kappa$ ), which is consistent with a singular isothermal elliptical (SIE) lens model. Although this simplification neglects external shear and more complex lens structures, it serves as a suitable approximation for this proof-of-principle study and does not impact our primary conclusions.

For the binary black hole (BBH) parameters, we randomly assign values for the component masses ( $m_1, m_2$ ),

inclination angle ( $\iota$ ), polarization angle ( $\psi$ ), right ascension ( $\alpha$ ), declination ( $\delta$ ), merger time ( $t_c$ ), and dimensionless spins ( $a_1, a_2$ ) according to the following probability distributions:

- Component masses are drawn from a power-law distribution with a peak, denoted as  $(m_1, m_2) \sim \text{power law} + \text{peak}$  [17].
- The inclination angle follows the distribution  $p(\iota) \propto \sin(\iota)$ , with  $\iota \in [0, \pi]$ .
- The polarization angle is uniformly distributed:  $p(\psi) \propto U(0, \pi)$ .
- The right ascension is uniformly distributed:  $p(\alpha) \propto U(0, 2\pi)$ .
- The declination follows the distribution  $p(\delta) \propto \cos(\delta)$ , with  $\delta \in [-\pi/2, \pi/2]$ .
- The merger time is uniformly distributed:  $p(t_c) \propto U(t_{\min}, t_{\max})$ , where the simulation spans 1 years.
- The dimensionless spin magnitudes are uniformly distributed between 0 and 0.99:  $p(a_1) \propto U(0, 0.99)$ .
- Similarly,  $p(a_2) \propto U(0, 0.99)$ .

For the simulation of the microlensing field, we follow the procedures outlined in Refs. [6, 18, 19]. We employ the Salpeter initial mass function [20] and an elliptical Sérsic profile [21] to model the stellar mass distribution and density associated with each strongly lensed gravitational wave (SLGW). The stellar mass range is set to [0.1, 1.5] solar masses, consistent with the value used by Diego *et al.* [22]. In addition to stars, we account for remnant objects within the microlensing field. We adopt the initial-final mass relation from Spera *et al.* [23] and assume that remnant objects contribute 20% to the total stellar mass density [8].

To calculate the frequency-dependent magnification caused by microlensing, we utilize the algorithm presented in Shan *et al.* [6], which involves evaluating the Fresnel-Kirchhoff diffraction integral [24]:

$$F(\omega, \mathbf{y}) = \frac{2GM_L(1+z_L)\omega}{\pi c^3 i} \int_{-\infty}^{\infty} d^2x \exp[i\omega t(\mathbf{x}, \mathbf{y})] , \quad (9)$$

where  $F(\omega, \mathbf{y})$  represents the wave optics magnification factor,  $\omega$  is the gravitational wave's circular frequency, and  $\mathbf{y}$  is its position in the source plane (normalized by the Einstein radius).  $M_L$  and  $z_L$  are the lens mass and redshift, respectively,  $\mathbf{x}$  denotes the lens plane coordinates, and  $t(\mathbf{x}, \mathbf{y})$  is the time-delay function, defined as:

$$t(\mathbf{x}, \mathbf{x}^i, \mathbf{y} = 0) = \underbrace{\frac{k}{2} ((1 - \kappa + \gamma)x_1^2 + (1 - \kappa - \gamma)x_2^2)}_{t_s(\kappa, \gamma, \mathbf{x})} - \underbrace{\left[ \frac{k}{2} \sum_i^N \ln (\mathbf{x}^i - \mathbf{x})^2 + k\phi_-(\mathbf{x}) \right]}_{t_m(\mathbf{x}, \mathbf{x}^i)} \quad (10)$$

Here,  $k = 4GM_{\text{micro}}(1+z_L)/c^3$  and  $\mathbf{x}^i$  represents the coordinate of the  $i$ th microlens.  $M_{\text{micro}}$  represents the average microlensing mass.

We define the macro image position as the coordinate origin ( $\mathbf{y} = 0$ ).  $\phi_-(\mathbf{x})$  represents the contribution from a negative mass sheet, introduced to compensate for the mass contribution from the microlenses and maintain a constant total convergence,  $\kappa$  [18, 19, 25]. The terms  $t_s(\kappa, \gamma, \mathbf{x})$  and  $t_m(\mathbf{x}, \mathbf{x}^i)$  represent the time delays due to the macro lens and the microlenses, respectively. With the completion of these steps, we have successfully generated all necessary components for our training data set.

### Test Data Set

To assess the performance of our methodology, we construct a test dataset following an astrophysical model, drawing inspiration from the methodologies presented in Refs. [13, 26]. The data generation process unfolds as follows:

- We generate BBH redshifts ( $z_s$ ) by sampling from a theoretical BBH merger rate model. This model assumes that the merger rate is proportional to the star formation rate (SFR) with a time delay of  $\Delta t = 50\text{Myr}$  between star formation and subsequent BBH formation. Refer to Appendix B of Xu *et al.* [13] for a detailed exposition of this model.
- For each simulated BBH event, we use the sampling distributions described in the training dataset section.
- For each BBH event at redshift  $z_s$ , we compute the multiple-image optical depth,  $\tau(z_s)$ , using the singular isothermal sphere (SIS) model, as described in Haris *et al.* [26]. We then draw a random number from a uniform distribution between 0 and 1. If the calculated optical depth exceeds the random number, we classify the event as a strongly lensed gravitational wave (SLGW) event; otherwise, it is excluded.
- For the selected SLGW events, we model the lensing galaxy as a singular isothermal ellipsoid

(SIE) [27] and utilize the Lenstronomy software package [28, 29] to solve the lens equation. The velocity dispersion ( $\sigma_v$ ) and axis ratio ( $q$ ) of the SIE are sampled from the observed distribution of the Sloan Digital Sky Survey (SDSS) galaxy population [30]. We note that Ref. [30] contains a typographical error in the axis ratio parameter, and we employ the corrected form as presented in Ref. [31]. The sampling details for these parameters, as well as the lens redshift and source-plane location, are detailed in Appendix A of Haris *et al.* [26].

We then use a detector network containing two Cosmic Explorer (CE) detectors, situated in Livingston (USA) and Hanford (USA).

Finally, for the simulation of the microlensing stellar field, we use the same method described in the training dataset section. Following the above procedure, we generate a test dataset of 100 events to evaluate the performance of the neural network.

### Normalizing flow and training

Given a micro-lensed waveform  $h_{lensed}$ , the goal of our model is to perform the density estimation  $q(\theta_{\text{GR}}, \theta_s | h_{lensed})$  of both BBH gravitational wave parameters  $\theta_{\text{GR}}$  and lensing parameters  $\theta_s$ . Following [11], we train a conditional normalizing flow – specifically a neural spline flows (NSF) [32] that models complex, potentially multimodal posteriors through adaptive spline transformations – which parameterizes a flexible probability distribution over the joint space of astrophysical and lensing parameters. Specifically, the flow learns an invertible transformation  $f_\phi$  that maps samples from a simple base distribution (e.g., a multivariate Gaussian) to samples from the complex, high-dimensional posterior distribution, conditioned on the observed lensed waveform data  $h_{lensed}$ . The parameters of the flow are optimized during training to maximize the likelihood of the training data under the model, effectively learning the density estimation of the full joint posterior distribution, which is critical for simultaneously inferring both the properties of the BBH system and the characteristics of the gravitational microlens.

wave event, Mon. Not. Roy. Astron. Soc. **497**, 204 (2020), arXiv:2007.00828 [astro-ph.CO].

- [3] T. Baker and M. Trodden, Multimessenger time delays from lensed gravitational waves, Phys. Rev. D **95**, 063512 (2017), arXiv:1612.02004 [astro-ph.CO].
- [4] T. E. Collett and D. Bacon, Testing the speed of gravitational waves over cosmological distances with strong gravitational lensing, Phys. Rev. Lett. **118**, 091101 (2017), arXiv:1602.05882 [astro-ph.HE].
- [5] X.-L. Fan, K. Liao, M. Biesiada, A. Piorkowska-Kurpas, and Z.-H. Zhu, Speed of Gravitational Waves from Strongly Lensed Gravitational Waves and Electromagnetic Signals, Phys. Rev. Lett. **118**, 091102 (2017), arXiv:1612.04095 [gr-qc].
- [6] X. Shan, G. Li, X. Chen, W. Zheng, and W. Zhao, Wave effect of gravitational waves intersected with a microlens field: a new algorithm and supplementary study, (2022), arXiv:2208.13566 [astro-ph.CO].
- [7] X. Shan, X. Chen, B. Hu, and G. Li, Microlensing bias on the detection of strong lensing gravitational wave, (2023), arXiv:2306.14796 [astro-ph.CO].
- [8] A. K. Meena, A. Mishra, A. More, S. Bose, and J. S. Bagla, Gravitational lensing of gravitational waves: Probability of microlensing in galaxy-scale lens population, Mon. Not. Roy. Astron. Soc. **517**, 872 (2022), arXiv:2205.05409 [astro-ph.GA].
- [9] J. M. Diego, O. A. Hannuksela, P. L. Kelly, T. Broadhurst, K. Kim, T. G. F. Li, G. F. Smoot, and G. Pagano, Observational signatures of microlensing in gravitational waves at LIGO/Virgo frequencies, Astron. Astrophys. **627**, A130 (2019), arXiv:1903.04513 [astro-ph.CO].
- [10] A. Nitz, I. Harry, D. Brown, C. M. Biwer, J. Willis, T. D. Canton, C. Capano, T. Dent, L. Pekowsky, A. R. Williamson, S. De, M. Cabero, B. Machenschalk, D. Macleod, P. Kumar, S. Reyes, dfinstad, F. Pannarale, S. Kumar, T. Massinger, M. Tápai, L. Singer, G. S. C. Davies, S. Khan, S. Fairhurst, A. Nielsen, S. Singh, K. Chandra, shasvath, and veronica villa, gwastro/pycbc: v2.0.2 release of pycbc (2022).
- [11] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, Real-time gravitational wave science with neural posterior estimation, Physical review letters **127**, 241103 (2021).
- [12] K. Wette, SWIGLAL: Python and Octave interfaces to the LALSuite gravitational-wave data analysis libraries, SoftwareX **12**, 100634 (2020).
- [13] F. Xu, J. M. Ezquiaga, and D. E. Holz, Please Repeat: Strong Lensing of Gravitational Waves as a Probe of Compact Binary and Galaxy Populations, Astrophys. J. **929**, 9 (2022), arXiv:2105.14390 [astro-ph.CO].
- [14] X. Shan, B. Hu, X. Chen, and R.-G. Cai, An interference-based method for the detection of strongly lensed gravitational waves, Nature Astron. **9**, 916 (2025), arXiv:2301.06117 [astro-ph.IM].
- [15] Z. Pan and H. Yang, Formation Rate of Extreme Mass Ratio Inspirals in Active Galactic Nuclei, Phys. Rev. D **103**, 103018 (2021), arXiv:2101.09146 [astro-ph.HE].
- [16] Z. Pan, Z. Lyu, and H. Yang, Wet extreme mass ratio inspirals may be more common for spaceborne gravitational wave detection, Phys. Rev. D **104**, 063007 (2021), arXiv:2104.01208 [astro-ph.HE].
- [17] B. P. Abbott *et al.* (LIGO Scientific, Virgo), Binary Black Hole Population Properties Inferred from the First and Second Observing Runs of Advanced LIGO and

\* hyangdoa@tsinghua.edu.cn

- [1] O. A. Hannuksela, T. E. Collett, M. Çalışkan, and T. G. F. Li, Localizing merging black holes with sub-arcsecond precision using gravitational-wave lensing, Mon. Not. Roy. Astron. Soc. **498**, 3395 (2020), arXiv:2004.13811 [astro-ph.HE].
- [2] H. Yu, P. Zhang, and F.-Y. Wang, Strong lensing as a giant telescope to localize the host galaxy of gravitational

- Advanced Virgo, *Astrophys. J. Lett.* **882**, L24 (2019), arXiv:1811.12940 [astro-ph.HE].
- [18] X. Chen, Y. Shu, G. Li, and W. Zheng, FRBs Lensed by Point Masses. II. The Multipeaked FRBs from the Point View of Microlensing, *Astrophys. J.* **923**, 117 (2021), arXiv:2110.07643 [astro-ph.GA].
- [19] W. Zheng, X. Chen, G. Li, and H.-z. Chen, An Improved GPU-based Ray-shooting Code for Gravitational Microlensing, *Astrophys. J.* **931**, 114 (2022), arXiv:2204.10871 [astro-ph.IM].
- [20] E. E. Salpeter, The Luminosity Function and Stellar Evolution., *Astrophys. J.* **121**, 161 (1955).
- [21] G. Vernardos, Microlensing flux ratio predictions for euclid, *Monthly Notices of the Royal Astronomical Society* **483**, 5583 (2018).
- [22] J. M. Diego, G. Bernstein, W. Chen, A. Goobar, J. P. Johansson, P. L. Kelly, E. Mörtzell, and J. W. Nightingale, Microlensing and the type Ia supernova iPTF16geu, *Astron. Astrophys.* **662**, A34 (2022), arXiv:2112.04524 [astro-ph.CO].
- [23] M. Spera, M. Mapelli, and A. Bressan, The mass spectrum of compact remnants from the PARSEC stellar evolution tracks, *451*, 4086 (2015), arXiv:1505.05201 [astro-ph.SR].
- [24] P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational Lenses* (Springer New York, NY, 1992).
- [25] J. Wambsganss, Ph.D. thesis, - (1990).
- [26] K. Haris, A. K. Mehta, S. Kumar, T. Venumadhav, and P. Ajith, Identifying strongly lensed gravitational wave signals from binary black hole mergers, (2018), arXiv:1807.07062 [gr-qc].
- [27] R. Kormann, P. Schneider, and M. Bartelmann, Isothermal elliptical gravitational lens models., *284*, 285 (1994).
- [28] S. Birrer and A. Amara, lenstronomy: Multi-purpose gravitational lens modelling software package, *Physics of the Dark Universe* **22**, 189 (2018), arXiv:1803.09746 [astro-ph.CO].
- [29] S. Birrer, A. Shajib, D. Gilman, A. Galan, J. Aalbers, M. Millon, R. Morgan, G. Pagano, J. Park, L. Teodori, N. Tessore, M. Ueland, L. Van de Vyvere, S. Wagner-Carena, E. Wempe, L. Yang, X. Ding, T. Schmidt, D. Sluse, M. Zhang, and A. Amara, lenstronomy II: A gravitational lensing software ecosystem, *The Journal of Open Source Software* **6**, 3283 (2021), arXiv:2106.05976 [astro-ph.CO].
- [30] T. E. Collett, The Population of Galaxy-Galaxy Strong Lenses in Forthcoming Optical Imaging Surveys, *Astrophys. J.* **811**, 20 (2015), arXiv:1507.02657 [astro-ph.CO].
- [31] A. R. A. C. Wierda, E. Wempe, O. A. Hannuksela, L. e. V. E. Koopmans, and C. Van Den Broeck, Beyond the Detector Horizon: Forecasting Gravitational-Wave Strong Lensing, *Astrophys. J.* **921**, 154 (2021), arXiv:2106.06303 [astro-ph.HE].
- [32] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, Neural spline flows, *Advances in neural information processing systems* **32** (2019).

## SUPPLEMENTARY MATERIALS

### Lensing parameter distribution in the training dataset

We present the distribution of the lensing parameters,  $l_p$  and  $l_d[M]$ , used in the training dataset. As shown in Fig. 6, all parameters exhibit distributions peaked around zero, consistent with the expectation that most gravitational-wave microlensing events induce only subtle distortions. The peak parameter in the time domain,  $l_p$ , shows a positive bias, which reflects its definition as the sum of peak values in the time domain. In contrast, the deviation parameters  $l_d[M]$  are centered around zero and display approximately symmetric distributions. This symmetry arises from the phase-based nature of the  $l_d[M]$  definition, which allows for both positive and negative fluctuations. In practice, these symmetric distributions contribute to well-calibrated  $p$ -values in flow-based posterior inference.

### Flowchart

We illustrate the overall workflow of the pipeline in Fig. 7. The training set consists of BBH signals generated using LALSimulation[12] and microlensed signals produced with our SLBBH Mock Data generation pipeline, described in the Methods section. The lensing parameters, including the peak value  $l_p$  and deviation values  $l_d[M]$ , are pre-computed during data generation. By uniformly sampling over  $\kappa$  and  $\kappa_s$  in the SLBBH Mock Data generation pipeline, the training set provides diverse microlensing conditions, allowing the network to learn waveform variations more effectively.

During training, we first extract a reduced basis from the BBH signals using singular value decomposition, which captures the dominant modes of waveform variation in the absence of microlensing. The full set of microlensed waveforms is then projected onto this basis to reduce dimensionality while retaining relevant physical features. In each training epoch, the BBH signals are shuffled and used once per epoch, while microlensed signals are randomly paired with BBH signals under the redshift constraint. On average, each microlensing waveform is reused approximately ten times per epoch. The training inputs are constructed by adding simulated noise drawn from the Cosmic Explorer power spectral density (with glitches removed). These inputs are passed to the Neural Spline Flows [32], which jointly estimates both the BBH source parameters and the microlensing parameters. After training, the modified Mahalanobis distance described in the Results section is used to quantify the confidence level of microlensing detection.

### Astrophysical Distribution of the 1000 samples

In the method section, we demonstrate the distribution of the modified Mahalanobis distance for 1000 samples under the astrophysical distribution. Here we show the sampled source BBH redshift distribution (Fig. 8 left) and the proportion of lensing and source redshift distribution (Fig. 8 right) of our 1000 injections.

### PE result for all predicted parameters

We show a typical PE result in Fig. 9, with all predicted parameters, including BBH parameters and microlensing parameters. As shown in Fig. 4, the orange and blue distributions correspond to the microlensed and the associated non-microlensed signals, respectively, as illustrated in the main paper “PE results and implications” part. Both injections yield consistent posterior distributions for the BBH parameters, as expected for gravitational-wave signals originating from the same binary system. In contrast, the lensing-related parameters exhibit distinct distributions (same as main paper Fig. 4), reflecting differences in their lensing configurations. This result demonstrates the capability of our framework to disentangle microlensing effects from strong lensing in a coherent and physically interpretable manner.

### PP-PLOT

To further evaluate our model’s ability for lensing parameter posterior estimation quantitatively, we perform a P–P plot analysis using 100 injection events sampled consistently from the training distribution, as shown in Fig. 10. For each event, we compute the percentile position of the true lensing parameter value within the corresponding marginalized posterior distribution and construct the empirical cumulative distribution function. The left panel of Fig. 10 shows the P–P plot restricted to microlensing parameters, while the right panel presents the result for all inferred parameters.

The Kolmogorov–Smirnov test yields  $p$ -values of 0.5893 for the lensing parameters and 0.1694 for the full parameter set, both suggesting no statistically significant deviation from the expected uniform distribution within the tested sample size. These results confirm that the posterior estimates produced by our model are well-calibrated and statistically consistent with the true injected values across the full parameter space.

### MICROLENSING SURROGATE PARAMETERS

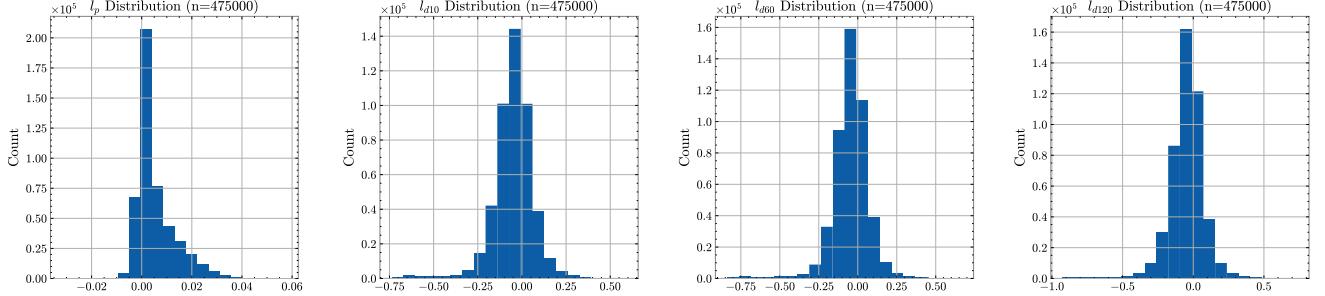


FIG. 6. Lensing parameter distribution in the training dataset.

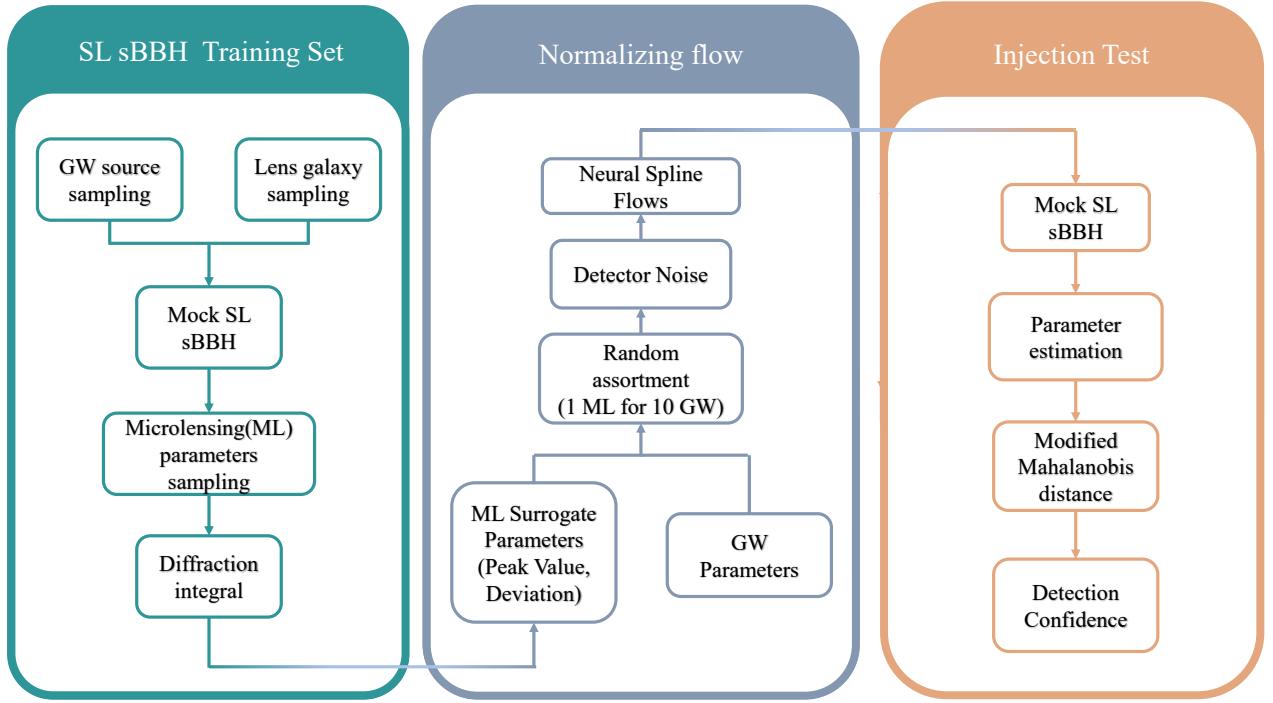


FIG. 7. Flowchart of our pipeline.

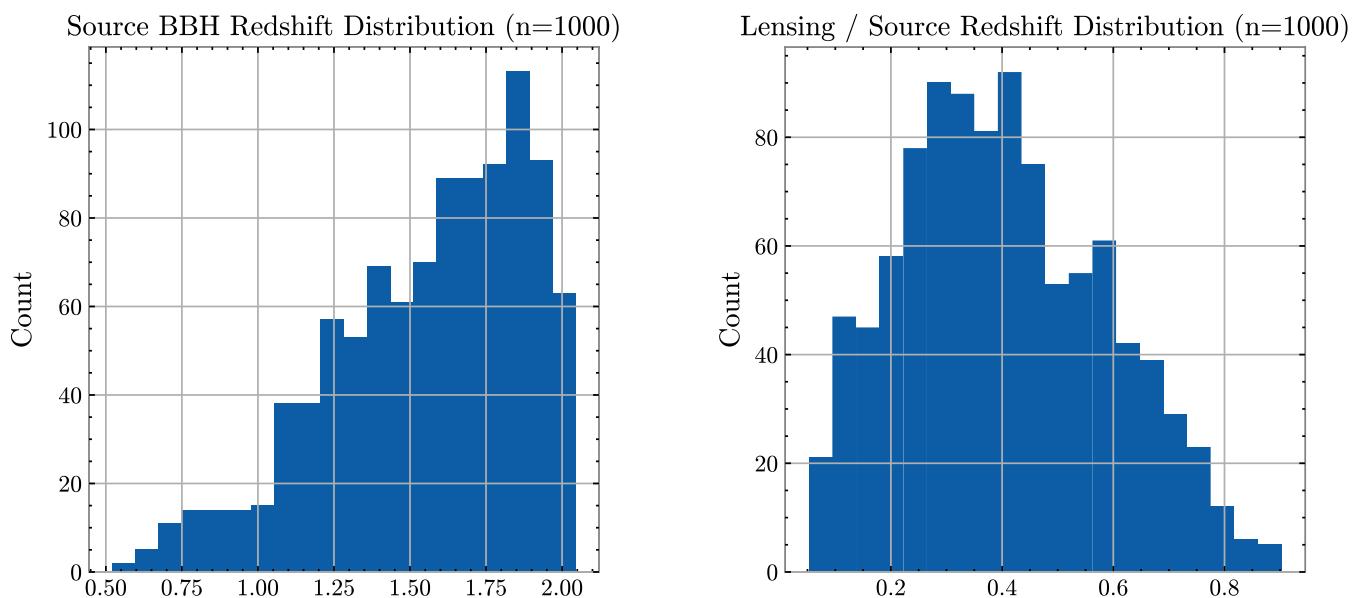


FIG. 8. Astrophysical distribution of the sampled 1000 injections.

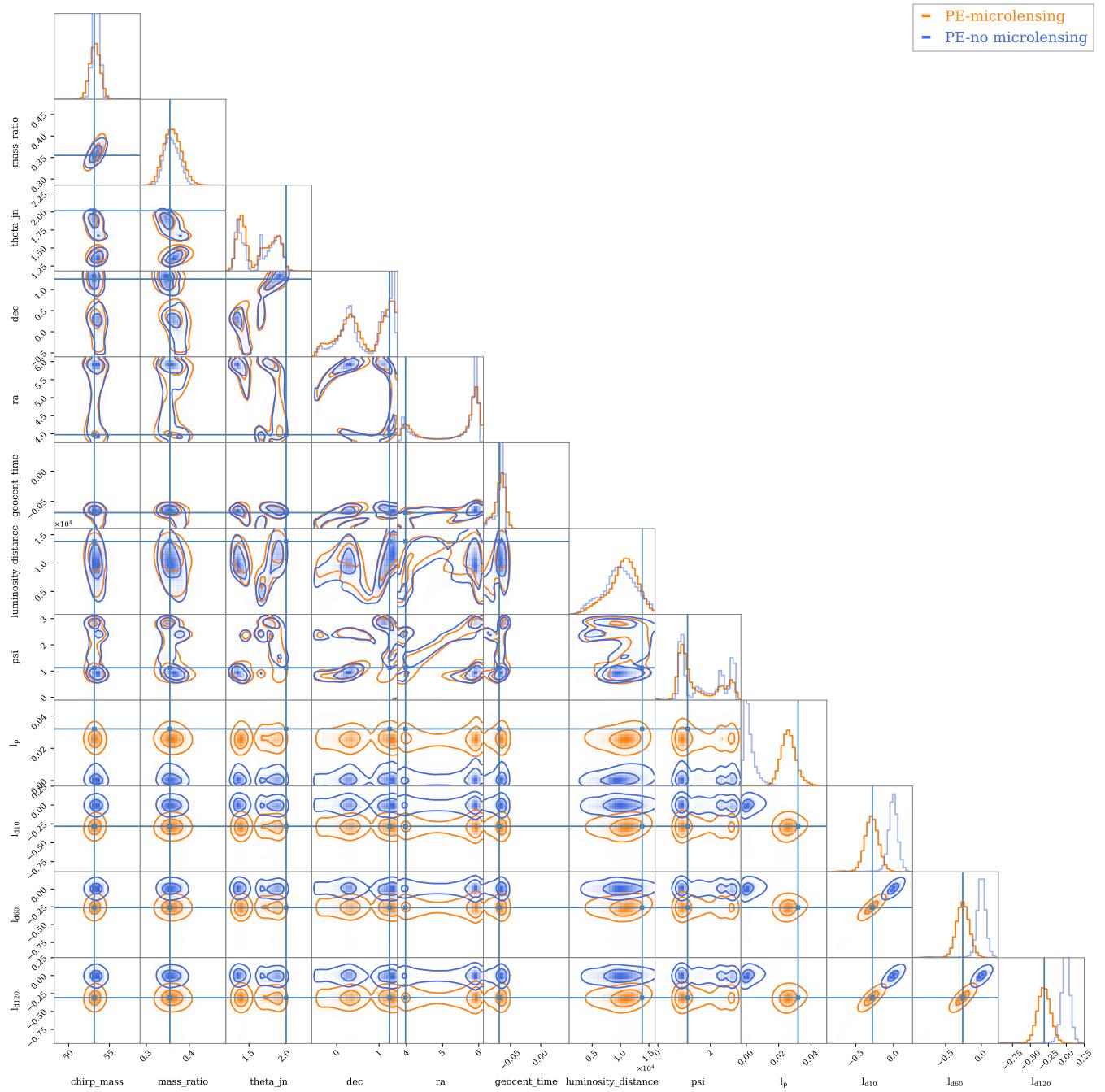


FIG. 9. PE result for all parameters.

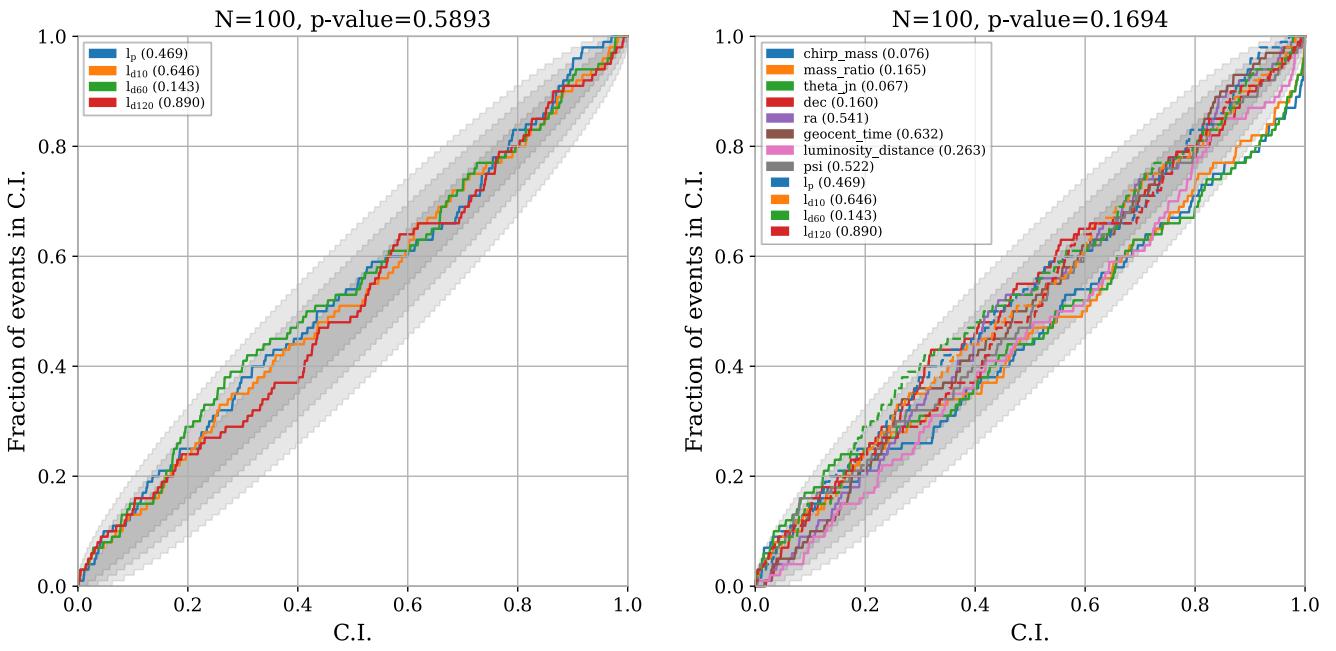


FIG. 10. P-P plot for all parameters.