UNIVERSITY OF COPENHAGEN
NIELS BOHR INSTITUTE

From Searches to Inference:

Towards the Discovery of Lensed Gravitational Waves

Doctoral Thesis

This thesis has been submitted to the PhD School of the Faculty of Science, University of Copenhagen.

To defend publicly in **Auditorium A, Niels Bohr Institute**
on Friday 27th February 2026 at 10.00

by

Juno Chun Lung Chan

Abstract

Gravitational waves (GWs) are ripples of spacetime generated by some of the most extreme systems in the Universe, such as coalescence of binary black holes. As GWs propagate across the Cosmos, intervening massive objects can bend their paths and distort their shapes, resulting the phenomenon known as *gravitational lensing*. The discovery of GW lensing would have far-reaching implications for both cosmology and GW astronomy, opening new avenues to probe compact object populations, the large scale structure of the Universe, and fundamental aspects of gravity. This thesis focuses on the unknowns and challenges associated with such a discovery across multiple stages of GW data analysis. We propose new strategies and develop a new method in matched-filtering searches to retrieve lensed GWs. We also address the computational bottleneck in statistical inference for GW lensing using deep-learning techniques.

We systematically investigate the limitations of matched-filtering searches in detecting lensed GWs and develop concrete improvements to enhance their detectability. The detectability of a GW signal is often assumed to be directly associated with its loudness. However, standard GW detection methods based on matched filtering employ waveform templates that neglect strong wave-optics effect. We show that a strong wave-optics effect can lead to a significant reduction in detection efficiency, even when the signal amplitude is boosted by an order of magnitude compared to the nonlensed signals. Our results reveal severe selection biases affecting multiple analyses, ranging from parameter inference to lensing rate estimation. To overcome these limitations, we propose new strategies for detecting highly distorted lensed GW signals across different detection methods. In parallel, we show that improvements in template bank construction and the development of dedicated detection statistics can enhance the detection of faint lensed GWs.

The high computational cost of lensing inference also hinders the discovery of GW lensing, since a large number of inferences is required to assess the statistical significance of a given lensing candidate. Building on top of a state-of-the-art simulation-based inference framework `DINGO`, we develop `DINGO-lensing` to overcome this limitation. It enables rapid inference of lensed GWs using deep learning while achieving accuracy comparable to traditional inference methods. We also develop inference strategies for discovering lensing for future LIGO-Virgo-KAGRA (LVK) observing runs and apply them to assess the statistical significance of GW231123, the event with the strongest support for lensing so far. This work marks an important milestone in the pursuit of GW lensing, transforming it from a computationally infeasible challenge into a statistically robust task with an affordable computational cost.

Altogether, we pave the way toward the first detection of GW lensing. We discuss the future prospects of

this line of research, highlighting how continued advances in detection and inference methods are steadily bringing this goal within reach of current ground-based detectors.

Resumé

Gravitationsbølger er rippel i rumtiden, der opstår fra nogle af Universets mest ekstreme systemer, såsom sammensmeltningen af to sorte huller. Når Gravitationsbølger bevæger sig gennem Kosmos, kan massive objekter på deres vej bøje deres baner og forvrænge deres former, hvilket producerer fænomenet kendt som *gravitationslinseeffekt*. Opdagelsen af gravitationslinseeffekt i Gravitationsbølger ville have vidtrækkende konsekvenser for både kosmologi og Gravitationsbølgeastronomi. Den ville åbne nye muligheder for at undersøge populationer af kompakte objekter, Universets storskala struktur og fundamentale aspekter af tyngdekraften.

Denne afhandling fokuserer på de ukendte faktorer og udfordringer, der er forbundet med en sådan opdagelse på flere trin i analysen af Gravitationsbølgdata. Vi foreslår nye strategier og udvikler en ny metode inden for matched filtering-søgninger for at kunne genfinde linsede Gravitationsbølger. Vi adresserer desuden den beregningsmæssige flaskehals i statistisk inferens for gravitationslinseeffekt i Gravitationsbølger ved hjælp af dyb læringsteknikker.

Vi undersøger systematisk begrænsningerne ved matched filtering-søgninger i detektion af linsede Gravitationsbølger og udvikler konkrete forbedringer for at øge deres detekterbarhed. Det antages ofte, at detekterbarheden af et Gravitationsbølgesignal er direkte relateret til dets styrke. Imidlertid benytter standardmetoder til detektion af Gravitationsbølger, baseret på matched filtering, bølgeformsskabeloner, der ser bort fra forvrængninger forårsaget af stærk linseeffekt. Vi viser, at en stærk linseforvrængning kan føre til en betydelig reduktion i detektionseffektiviteten, selv når signalets amplitude forstærkes med en størrelsesorden sammenlignet med ikke-linsede signaler. Vores resultater afslører alvorlige selektionsskævheder, der påvirker flere analyser, lige fra parameterestimation til beregning af forekomsten af linseeffekt. For at overvinde disse begrænsninger foreslår vi nye strategier til at opdage stærkt forvrængede Gravitationsbølgesignaler forårsaget af linseeffekt på tværs af forskellige detektionsmetoder. Parallelt viser vi, at forbedringer i udformningen af skabelonbanker og udviklingen af dedikerede detektionsstatistikker kan øge chancen for at opdage svage linsede Gravitationsbølger.

Den høje beregningsmæssige omkostning ved linseinferens hæmmer også opdagelsen af gravitationslinseeffekt i Gravitationsbølger, da et stort antal inferenser kræves for at vurdere den statistiske signifikans af en given kandidat til linseeffekt. Med udgangspunkt i en avanceret, simuleringsbaseret inferensramme `DINGO` udvikler vi `DINGO-lensing` for at overvinde denne begrænsning. Den muliggør hurtig inferens af linsede Gravitationsbølger ved hjælp af dyb læring, samtidig med at den opnår nøjagtighed, der kan sammenlignes med traditionelle inferensmetoder. Vi udvikler desuden inferensstrategier til at opdage linseeffekt i fremtidige observationsrunder af Gravitationsbølgedetektorer og anvender dem til at vurdere

den statistiske signifikans af hændelsen GW231123, som indtil nu har den stærkeste støtte for at være påvirket af linseeffekt. Dette arbejde markerer en vigtig milepæl i jagten på linseeffekt i Gravitationsbølger og gør det muligt at omdanne et tidligere beregningsmæssigt uoverkommeligt problem til en statistisk robust opgave med overkommelig beregningsomkostning.

Alt i alt baner vi vejen for den første detektion af linseeffekt i Gravitationsbølger. Vi diskuterer de fremtidige udsigter for denne forskningsretning og fremhæver, hvordan fortsatte fremskridt i detektion og inferensmetoder gradvist bringer dette mål inden for rækkevidde af de nuværende jordbaserede detektorer.

Acknowledgments

It has been an exciting journey towards the preparation of this thesis. The scientific growth would be impossible without the passionate and excellent people around me.

I would like to thank my supervisor, *Jose María Ezquiaga*, for his constant support, guidance, and insightful advice throughout this work. I am also grateful to my supervisor, *Vitor Cardoso*, for the constructive discussions and for his continued support during my doctoral studies. Both of you have been exceptional role models, and your mentorship has had a lasting influence on my development as a researcher.

I would also like to thank all my *colleagues* in the strong gravity group for creating a stimulating and supportive research environment, and for the inspiration drawn from your dedication to physics. Special thanks go to all my *collaborators*, whose insights and expertise have broadened the scope and possibilities of my research and have greatly shaped the scientific outcomes of this work. All of you shaped my academic and personal growth.

I would like to thank my sister, *Ashley*, and my family for all their emotional support. The cover page artwork is credited to *Tiffany*, who joined this journey midway, and to whom I am deeply grateful for the patience and encouragement that made this path both enriching and emotionally sustaining. Thank you all.

Publications

This thesis is based on a series of publications that investigate lensed gravitational wave (GW) as listed below and are sorted by the time of publication.

1. **Juno C. L. Chan** et al., *Discovering gravitational waveform distortions from lensing: a deep dive into GW231123*, arXiv:2512.16916, (Submitted to Phyical Review Letter)

2. **Juno C. L. Chan** et al., *Identification and characterization of distorted gravitational waves by lensing using deep learning*, Phys. Rev. D 113. 024041, (2025);

3. **Juno C. L. Chan** et al., *Detectability of lensed gravitational waves in matched-filtering searches*, Phys. Rev. D 111.084019, (2025)

4. Alvin K. Y. Li, **Juno C. L. Chan** et al., *TESLA-X: An effective method to search for sub-threshold lensed gravitational waves with a targeted population model*, MNRAS 542, 998-1010, (2025)

Chapter 2 focuses on the detectability of lensed GWs and the impact of lensing induced waveform distortions on standard search pipelines [1]. Chapter 3 is based on the development of `DINGO-lensing` framework to enable efficient simulation-based inference for lensed GWs [2]. These methods are subsequently applied to the analysis of the high ranking lensing candidate GW231123 using `DINGO-lensing` [3]. Chapter 4 incorporates contributions to TESLA-X, a complementary search framework designed to enhance sensitivity to demagnified GW signals [4].

I also authored the following publications during my doctoral studies, which are not included in this thesis. The publications are acknowledged and are sorted by the time of publication as listed below.

1. LIGO Scientific and VIRGO and KAGRA Collaborations, A. G. Abac, ..., **Juno C. L. Chan**, ..., *GWTC-4.0: Searches for Gravitational-Wave Lensing Signatures*, arXiv:2512.16347 [5]

2. **Juno C. L. Chan**, Conor Dyson, Matilde Garcia, Jaime Redondo-Yuste, Luka Vujeva, *Lensing and wave optics in the strong field of a black hole*, Phys. Rev. D 112, 064009, (2025) [6]

3. Mick Wright, Justin Janquart, Paolo Cremonese, **Juno C. L. Chan** et al., *LensingFlow: An Automated Workflow for Gravitational Wave Lensing Analyses*, arXiv:2507.20256, (2025), (Accepted by RAS Techniques and Instruments) [7]

4. Graham P. Smith, ..., **Juno C. L. Chan**, ..., *Multi-messenger gravitational lensing*, Phil. Trans. Roy. Soc. Lond.A 383 (2025) 2295, 20240134, (2025) [8]

5. Luka Vujeva, Rico K. L. Lo, Jose M. Ezquiaga, **Juno C. L. Chan**, *Effects of Galaxy Cluster Structure on Lensed Transients*, Phys. Rev. D 112, 063044, (2025) [9]

6. **Juno C. L. Chan** et al., *Distinct neutrino signatures of quark deconfinement in accretion-induced collapse of white dwarfs*, arXiv:2412.10046 [10];

7. Patrick Chi-Kit Cheong, Antonios Tsokaros, Milton Ruiz, Fabrizio Venturi, **Juno C. L. Chan** et al., *General-relativistic resistive-magnetohydrodynamics simulations of self-consistent magnetized rotating neutron stars*, Phys. Rev. D 111.063030, (2025) [11]

8. Justin Janquart, David Keitel, Rico K.L. Lo, **Juno C. L. Chan** et al., *What is the nature of GW230529? An exploration of the gravitational lensing hypothesis*, MNRAS 537, 1001–1014, (2025) [12]

9. Rico K. L. Lo, Luka Vujeva, Jose M. Ezquiaga, **Juno C. L. Chan**, *Observational Signatures of Highly Magnified Gravitational Waves from Compact Binary Coalescence*, Phys. Rev. Lett. 134.151401, (2025) [13]

10. Luka Vujeva, Rico K. L. Lo, Jose M. Ezquiaga, **Juno C. L. Chan**, *lenscat: a public and community-contributed catalogue of known strong gravitational lenses*, Phil. Trans. Roy. Soc. Lond.A 383 (2025) 2294, 20240168, (2025) [14]

11. Srashti Goyal, Shasvath J. Kapadia, Jean-Rene Cudell, Alvin K. Y. Li, **Juno C. L. Chan**, *Rapid method for preliminary identification of subthreshold strongly lensed counterparts to superthreshold gravitational-wave events*, Phys. Rev. D 109.023028, (2024) [15]

12. LIGO Scientific and Virgo Collaborations, R. Abbott, ..., **Juno C. L. Chan**, ..., *Search for Gravitational-lensing Signatures in the Full Third Observing Run of the LIGO–Virgo Network*, Astrophys. J. 923.2, (2024) [16]

13. Justin Janquart, Mick Wright, Srashti Goyal, **Juno C. L. Chan** et al., *Follow-up analyses to the O3 LIGO–Virgo–KAGRA lensing searches*, MNRAS 526, 3832–3860 , (2023) [17]

14. Alvin K. Y. Li, Rico K. L. Lo, Surabhi Sachdev, **Juno C. L. Chan** et al., *Targeted sub-threshold search for strongly lensed gravitational-wave events*, Phys. Rev. D 107.123014, (2023) [18]

15. LIGO Scientific and Virgo Collaborations, R. Abbott, ..., **Juno C. L. Chan**, ..., *Search for subsolar-mass black hole binaries in the second part of Advanced LIGO's and Advanced Virgo's third observing run*, MNRAS 524, 5984–5992, (2023) [19]

16. LIGO Scientific and Virgo Collaborations, R. Abbott, ..., **Juno C. L. Chan**, ..., *Search for Lensing Signatures in the Gravitational-Wave Observations from the First Half of LIGO–Virgo's Third Observing Run* Astrophys. J. 923.1, (2021) [20]

Conventions and Units

Unless otherwise stated, we work in natural units, for which the speed of light c, gravitational constant G, solar mass $M_\odot$, vacuum permittivity ϵ_0 and vacuum permeability μ_0 are all equal to one ($c = G = M_\odot = \epsilon_0 = \mu_0 = 1$). Greek indices $\alpha, \beta, \delta, \cdots$, running from 0 to 3, are used for 4-quantities while the Roman indices such as $i, j, k, \cdots$, running from 1 to 3, are used for 3-quantities. Throughout this thesis, we adopt the $(-,+,+,+)$ metric signature. Therefore, the metric tensor for flat spacetime in (t, x, y, z) coordinates, is given by

$$\eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1). \tag{0.1}$$

The spacetime interval can then be constructed by

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu. \tag{0.2}$$

A comma followed by an index gives a partial derivative,

$$V^\nu{}_{,\mu} \equiv \frac{\partial V^\nu}{\partial x^\mu}, \tag{0.3}$$

A semi-colon followed by an index gives a covariant derivative,

$$V^\nu{}_{;\mu} \equiv V^\nu{}_{,\mu} + V^\alpha \Gamma^\nu{}_{\alpha\mu}. \tag{0.4}$$

where the Christoffel symbol $\Gamma^\rho_{\mu\nu}$ is given by

$$\Gamma^\rho{}_{\mu\nu} = \frac{1}{2} g^{\sigma\rho} (g_{\sigma\mu,\nu} + g_{\sigma\nu,\mu} - g_{\mu\nu,\sigma}). \tag{0.5}$$

The Riemann tensor is given by

$$R^\mu{}_{\nu\rho\sigma} \equiv \Gamma^\mu{}_{\nu\sigma,\rho} - \Gamma^\mu{}_{\nu\rho,\sigma} + \Gamma^\mu{}_{\alpha\rho}\Gamma^\alpha{}_{\nu\sigma} - \Gamma^\mu{}_{\alpha\sigma}\Gamma^\alpha{}_{\nu\rho}. \tag{0.6}$$

The Ricci tensor and Ricci scalar are defined by

$$R_{\mu\nu} \equiv R^\alpha{}_{\mu\alpha\nu}, \text{ and } R \equiv g^{\mu\nu} R_{\mu\nu} \tag{0.7}$$

Contents

Chapter 1

Introduction

In this chapter, we introduce the core concepts of gravitational waves (GWs) that frame the discussions in the chapters to come [21]. Section 1.1 begins with the essentials: how GWs are generated and how compact binary mergers produce the signals we observe [21]. Section 1.2 then turns to the detectors themselves, describing how matched-filtering search pipelines extract signals from noisy data [22, 23]. Section 1.3 explains how significant events are analysed through Bayesian inference [24]. Finally, Section 1.4 presents the theory of gravitational lensing and its observational impact on the waveforms, and offers an overview of current efforts to identify these lensing signatures [25, 26].

1.1 Basics of gravitational waves

In general relativity, gravity is no longer viewed as a force in the Newtonian sense but as a manifestation of the curvature of spacetime itself. This curvature is determined by the Einstein field equations,

$$G^{\mu\nu} = 8\pi T^{\mu\nu}, \tag{1.1}$$

in which the Einstein tensor $G^{\mu\nu}$ encodes how spacetime is curved and is built from the Ricci tensor $R^{\mu\nu}$, the Ricci scalar R, and the metric $g^{\mu\nu}$,

$$G^{\mu\nu} \equiv R^{\mu\nu} - \frac{1}{2} g^{\mu\nu} R. \tag{1.2}$$

On the right hand side, the stress–energy tensor $T^{\mu\nu}$ describes the content of the universe: how energy, momentum, and pressure are distributed and flow through spacetime.

Taken together, these equations form a set of ten coupled, nonlinear differential equations that are notoriously difficult to solve exactly. In many situations, progress is only possible by simplifying the problem: for example, in regions where gravity is weak, one can linearize the equations, reducing them to a wave equation that describes small perturbations of flat spacetime. In more extreme regimes, such as merging black holes, one must instead solve the full nonlinear equations numerically using specialized formulations. For now, the discussion is restricted to the linearized theory of general relativity, which

captures many essential features of GW.

1.1.1 Linearized theory of general relativity

The basic features of GWs can be brought out cleanly by working in the weak field regime, where gravity is treated as a small perturbation of flat spacetime. In this approximation, only the lowest order deviations from the Minkowski metric are retained, and the Einstein equations reduce to a linear theory that is much easier to analyze. The remainder of this section will develop the standard linearized form of the field equations and use it to explore the main properties of GWs, including how they propagate on flat spacetime, how they act on freely falling test masses, and how they are generated by astrophysical sources.

1.1.1.1 Weak-field limit

The Einstein field equations are covariant under arbitrary coordinate transformations, meaning their form is preserved in any spacetime coordinate system. In the weak field regime, one assumes that spacetime is only slightly curved so that the metric can be written as a small perturbation of flat Minkowski space,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{1.3}$$

where the perturbation satisfies $|h_{\mu\nu}| \ll 1$. However, in the weak-field limit the full general coordinate invariance is effectively reduced, so not all coordinate transformations are allowed, and one must impose additional conditions on the coordinates.

For a gauge transformation, consider

$$x^{\mu} \to x^{\mu'} = \eta^{\mu'}_{\ \nu}(x^{\nu} + \xi^{\nu}), \tag{1.4}$$

where the gauge vector satisfies $|\xi_{\mu,\nu}| \lesssim |h_{\mu\nu}|$. Keeping only terms linear in the perturbation, the metric transforms as

$$h_{\mu\nu} \to h_{\mu'\nu'} = h_{\mu\nu} - \xi_{\mu,\nu} - \xi_{\nu,\mu}, \tag{1.5}$$

under which the Einstein field equations in the weak field limit remain invariant.

In addition, it is straightforward to verify that the Lorentz transformation law for $g_{\mu\nu}$ induces the same transformation on the perturbation $h_{\mu\nu}$,

$$h_{\mu\nu} \to h_{\mu'\nu'} = \Lambda^{\rho}_{\ \mu'}\Lambda^{\sigma}_{\ \nu'}h_{\rho\sigma}, \tag{1.6}$$

where $\Lambda^{\rho}_{\ \mu'}$ denotes the Lorentz transformation matrix.

1.1.1.2 Linearized Einstein field equations

Keeping only the leading order terms in $h_{\mu\nu}$, the Einstein field equations reduce to the linearized form

$$16\pi T_{\mu\nu} = \bar{h}_{\mu\alpha}{}^{,\alpha}{}_{\nu} + \bar{h}_{\nu\alpha}{}^{,\alpha}{}_{\mu} - \bar{h}_{\mu\nu,\alpha}{}^{\alpha} - \eta_{\mu\nu}\bar{h}_{\alpha\beta}{}^{,\alpha\beta}, \tag{1.7}$$

where the trace of the perturbation,

$$h \equiv \eta^{\alpha\beta}h_{\alpha\beta}, \tag{1.8}$$

is used to define the trace–reversed field

$$\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h. \tag{1.9}$$

A particularly convenient choice is the harmonic gauge (also known as Lorenz gauge and de Donder gauge),

$$\bar{h}_{\mu\nu}{}^{,\nu} = 0 \tag{1.10}$$

under which the linearized equations simplify dramatically to

$$\bar{h}_{\mu\nu,\alpha}{}^{\alpha} = -16\pi T_{\mu\nu}, \tag{1.11}$$

a wave equation that forms the starting point for studying generation properties of GWs (See Section 1.1.4). The harmonic gauge imposes four conditions on the symmetric metric perturbation, which originally has ten independent components, and preserves the Lorentz invariance of the wave equation. Even so, the gauge is not yet completely fixed: using the gauge transformation in Eq. (1.5), one can impose four additional constraints by requiring

$$\xi_{\mu,\nu}{}^{\nu} = 0. \tag{1.12}$$

Therefore, imposing harmonic gauge (Eq. (1.10)) and gauge condition in Eq. 1.12 has reduced the degree of freedom of $g_{\mu\nu}$ into two.

Together, the harmonic gauge condition and this residual gauge fixing reduce the physical degrees of freedom of $g_{\mu\nu}$ down to two, corresponding to the two independent polarizations of GWs.

1.1.2 Propagation of gravitational waves

To study the propagation of GWs, we consider linear metric perturbations $h_{\mu\nu}$ on a fixed background spacetime with metric $g^{(B)}_{\mu\nu}$, assuming that the background satisfies the Einstein field equations, typically in vacuum. In this framework, one works in the harmonic gauge, where the trace-reversed perturbation

$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2} g^{(B)}_{\mu\nu} h$ satisfies the gauge condition in Eq. 1.10.

When appropriate, for example in asymptotically flat regions, one may further impose transverse and traceless conditions in order to isolate the physical radiative degrees of freedom (See Section 1.1.3).

Under these assumptions, the propagation of GWs on a curved background spacetime is governed by the curvature coupled wave equation

$$h_{\mu\nu}{}^{;\alpha}_{;\alpha} + 2R^{(B)}_{\alpha\mu\beta\nu} h^{\alpha\beta} = 0, \tag{1.13}$$

where $R^{(B)}_{\alpha\mu\beta\nu}$ is the background Riemann tensor with respecto the background metric $g^{(B)}_{\mu\nu}$. The second term explicitly encodes the coupling between the GW and the curvature of the background spacetime.

Since Eq. (1.13) is linear in $h_{\mu\nu}$, the superposition principle applies, and solutions may be decomposed into elementary wave modes. In regions where the background curvature varies slowly compared to the GW wavelength, the equation admits approximate plane wave solutions, which form the basis for geometric optics and WKB type descriptions of GW propagation.

$$\bar{h}_{\mu\nu} = \mathbb{R}(A_{\mu\nu} e^{i k_\alpha x^\alpha}), \tag{1.14}$$

where $\mathbb{R}$ denotes the real part, $A_{\mu\nu}$ is a constant polarization tensor, and k_α is the wave four vector. Substituting this ansatz into the wave equation shows that

$$k_\alpha k^\alpha = 0, \tag{1.15}$$

so the waves propagate along the light cone. The harmonic gauge condition further implies

$$A_{\mu\nu} k^\nu = 0, \tag{1.16}$$

which means that the perturbation is transverse to the direction of propagation; GWs are therefore transverse waves.

1.1.3 Transverse-traceless gauge

There is still residual gauge freedom that can be used to more fully specify $A_{\mu\nu}$. One may impose the condition

$$A_{\mu\nu} u^\nu = 0, \tag{1.17}$$

where u^μ is a fixed four velocity, representing a congruence of observers filling spacetime. This condition effectively supplies three independent constraints on $A_{\mu\nu}$, since one combination is redundant due to $k^\mu (A_{\mu\nu} u^\nu) = 0$. A fourth condition can be added by requiring the perturbation to be traceless,

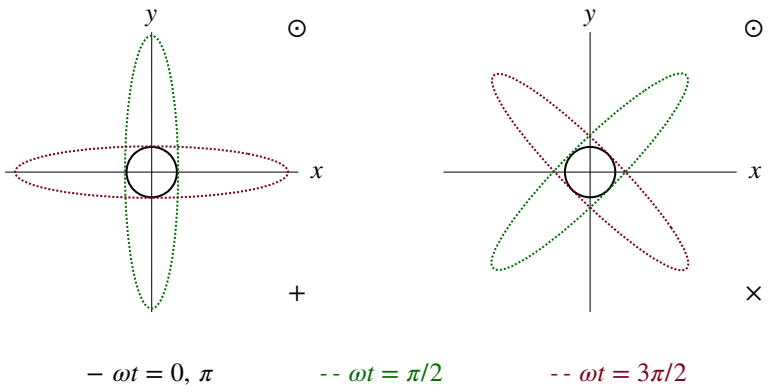

Figure 1.1: Response of a circular ring of test particles to the passage of a GW. The plus polarization distorts (+) the ring along horizontal and vertical axes, while the cross polarization (×) shears it along diagonals. The deformation patterns shown in different colours correspond to successive phases of the wave cycle, illustrating the characteristic quadrupolar nature of GW polarization.

$$A_{\nu}^{\ \nu} = 0. \tag{1.18}$$

In the comoving frame where $u^0 = 1$ and $u^i = 0$, these requirements translate into the transverse–traceless (TT) gauge conditions on the metric perturbation,

$$\begin{aligned} \bar{h}_{\ j}^{\,j} &= 0, \\ \bar{h}_{\mu 0} &= 0, \\ \bar{h}_{ij,}^{\ \ \ j} &= 0. \end{aligned} \tag{1.19}$$

A convenient feature of the TT gauge is that $h_{\mu\nu} = \bar{h}_{\mu\nu}$, so one can drop the bar. For a plane wave propagating in the z direction, the TT gauge metric perturbation can then be written as

$$h_{\mu\nu}^{TT} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cos[\omega(t-z)]. \tag{1.20}$$

where h_+ and $h_\times$ describe the two independent polarization states of the GW.

Quantities expressed in the transverse–traceless gauge are indicated by the superscript TT. The resulting deformations of freely falling test masses under the action of the plus and cross GW polarizations are illustrated in Fig. 1.1, where each polarization produces its characteristic stretching and squeezing pattern in the plane orthogonal to the direction of propagation.

1.1.4 Generation of gravitational waves

The generation of GWs is described by the wave equation given in Eq. 1.11. This equation admits a solution obtained using the Green's function approach. Expressed in terms of the retarded time and the spatial coordinates, the metric perturbation at the field point is

$$\bar{h}_{\mu\nu}(t, \vec{x}) = 4 \int d^3\vec{y}\, \frac{T_{\mu\nu}(t - |\vec{x} - \vec{y}|, \vec{y})}{|\vec{x} - \vec{y}|}, \tag{1.21}$$

The integral shows that the gravitational field at time t and position $\vec{x}$ is determined by the stress energy tensor evaluated at earlier times along the past light cone. The integration variable $\vec{y}$ spans only the spatial region occupied by the source, reflecting the assumption of a spatially localized emitter. To extract the physical GW degrees of freedom, the solution in Eq. (1.21) is projected onto the transverse traceless frame. This is achieved by introducing the TT projection operator

$$\Lambda_{ijkl} = P_{ik}P_{jl} - \frac{1}{2}P_{ij}P_{kl}. \tag{1.22}$$

Here both longitudinal components and the trace of the metric perturbation are removed. Applying this operator to the spatial components of the trace reversed metric yields the transverse traceless metric perturbation

$$\begin{aligned} h^{\mathrm{TT}}_{ij}(t, \vec{x}) &= \Lambda_{ijkl}\, \bar{h}_{kl} \\ &= 4\, \Lambda_{ijkl} \int d^3\vec{y}\, \frac{T_{kl}(t - |\vec{x} - \vec{y}|, \vec{y})}{|\vec{x} - \vec{y}|} \end{aligned} \tag{1.23}$$

For an isolated, far away, non-relativistic source, the expression can be further simplified. Using multipole expansion, the zeroth moment and dipole moment are time-independent due to conservation of energy and momentum respectively. Therefore, the leading contribution to GW radiation in a system is the quadrupole moment

$$M^{ij} = \int d^3\vec{y}\, T^{00}(t, \vec{y})\, y^i y^j, \tag{1.24}$$

in which the second time derivative lead to GW radiation

$$\left[h^{\mathrm{TT}}_{ij}(t, \vec{x})\right]_{\mathrm{quad}} = \frac{2}{r}\, \Lambda_{ijkl}\, \ddot{M}^{kl}(u). \tag{1.25}$$

1.1.4.1 Compact binary coalescences

The most important sources for ground-based detectors are compact binary coalescences, in which two compact objects, typically black holes or neutron stars, orbit each other and gradually inspiral as they lose energy to gravitational radiation. In the simplest description, a quasi circular binary with component

masses m_1 and m_2 is characterized by its total mass $M = m_1 + m_2$, mass ratio $q = m_2/m_1 \leq 1$, and chirp mass in source frame

$$\mathcal{M}_c = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}. \tag{1.26}$$

which controls the leading order phase evolution of the signal [27, 28]. At quadrupole order, the GW frequency is twice the orbital frequency in the detector frame, $f_{\rm GW} \simeq 2 f_{\rm orb}$ .In particular, for a circular orbit, the strain amplitude including the effects of cosmological propagation scales as

$$h(t) \propto \frac{\mathcal{M}_c^{5/3}}{d_L} \left[\pi f_{\rm GW}(t)\right]^{2/3}, \tag{1.27}$$

where d_L is the luminosity distance to the source. As the system inspirals, $f_{\rm GW}(t)$ and the amplitude grow according to the "chirp" law

$$\frac{d f_{\rm GW}}{dt} \propto \mathcal{M}_c^{5/3} \left[\pi f_{\rm GW}(t)\right]^{11/3}, \tag{1.28}$$

so the signal sweeps upward in frequency and amplitude until the binary merges and settles to a final compact remnant.

The (dimensionless) spin of each component is defined as

$$\vec{\chi}_i = \frac{c \vec{S}_i}{G m_i^2}, \tag{1.29}$$

The individual dimensionless spin parameters can be combined into a single effective inspiral spin parameter, $\chi_{\rm eff}$ [29, 30], which captures the leading order spin effects in the post Newtonian expansion and is defined as

$$\chi_{\rm eff} = \frac{(m_1 \vec{\chi}_1 + m_2 \vec{\chi}_2) \cdot \hat{L}_{\rm N}}{m_1 + m_2}, \tag{1.30}$$

where $\hat{L}_{\rm N}$ is the unit vector along the Newtonian orbital angular momentum.

Figure 1.2 shows a representative time-domain GW strain signal from a compact binary coalescence observed by a ground-based detector. Both amplitude and instantaneous frequency rise as the binary loses orbital energy through gravitational radiation. At early times, when the binary components are widely separated, the signal oscillates slowly with a relatively small amplitude. As the inspiral progresses, the GW frequency increases following the chirp law in Eq. 1.28, while the strain amplitude grows according to Eq. 1.27. This evolution highlights the central role of $\mathcal{M}_c$ in determining the leading-rder phase and amplitude behaviour throughout the inspiral. In the final stages, the sharp rise in frequency and amplitude signals the approach to merger, after which the waveform rapidly terminates as the system forms a single compact remnant. This overall signal structure is universal among compact binary coalescences and forms

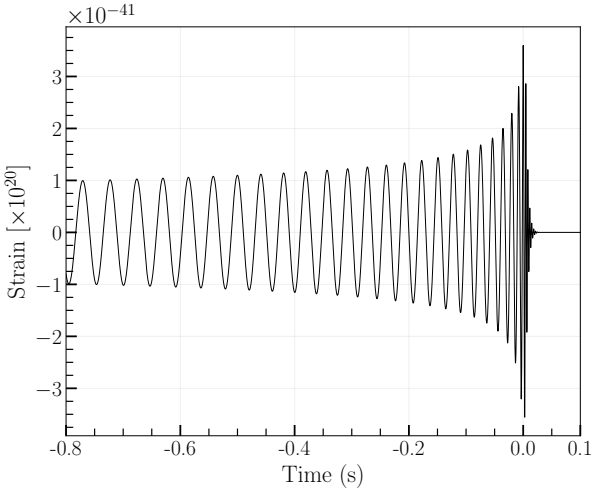

Figure 1.2: Time domain GW strain from a compact binary coalescence, illustrating the characteristic chirp signal. The amplitude and frequency increase as the binary inspirals due to energy loss to gravitational radiation, culminating in the merger of the two compact objects.

the basis of matched-filtering searches in GW detection.

As GW propagate through cosmic distance, the expansion of the Universe rescale the GW amplitude and produce a redshift in the observed frequency. The GW luminosity distance is

$$d_L = (1+z)\int_0^z \frac{c}{H(z)}\,dz, \tag{1.31}$$

where c is the speed of light and z is the redshift. Also, the chirp mass in the source frame and the detector frame is related by

$$\mathcal{M}_c(1+z) = \mathcal{M}_z. \tag{1.32}$$

In the following chapters, the chirp mass is defined in the detector frame unless specified otherwise.

1.2 Extracting signals from noise: matched-filtering searches

In Section 1.1, we introduced the fundamental physics of GWs, including their governing wave equations, generation mechanisms, and propagation effects within general relativity. In this chapter, we shift focus to the experimental and data analysis aspects of GW astronomy. We present the essential principles underlying GW detectors, describe how these instruments respond to passing GWs, and explain the methods used to extract weak GW signals from noisy detector data [21, 22].

1.2.1 Basics of gravitational-wave detection

As illustrated in Fig. 1.2, the plus and cross polarizations of a GW act to alternately stretch and squeeze the proper separations between freely falling test particles in orthogonal directions. This periodic deformation

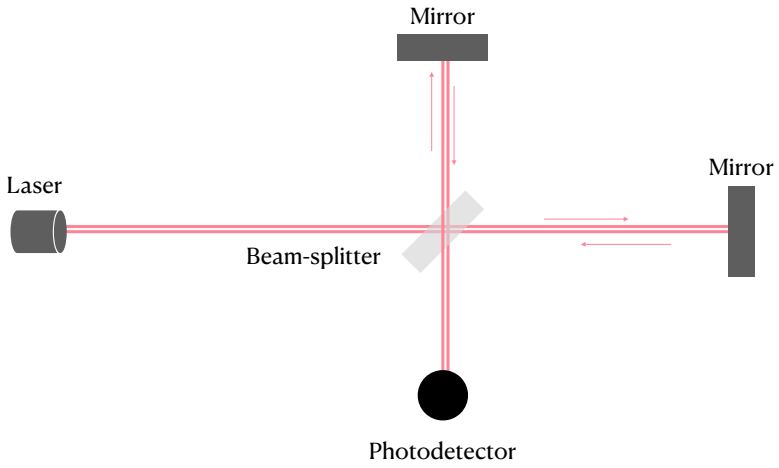

Figure 1.3: Schematic layout of a Michelson interferometric GW detector. A laser beam is split at the beam splitter into two orthogonal arms, reflected by end mirrors, and recombined at the beam splitter. In the absence of a GW the beams interfere destructively at the output port, while a passing GW induces differential arm length changes that produce a measurable signal at the photodetector.

modifies the proper distances between the particles.

In what follows, we focus on interferometric detectors, which are designed to measure these tiny GW induced changes in proper distance. An interferometer is configured such that, in the absence of a GW, the light beams traveling along its orthogonal arms recombine destructively at the output port, producing strongly reduced signal (See Figure 1.3). When a GW passes through the detector, it differentially alters the arm lengths, changing the optical path lengths traversed by the laser beams. This perturbation breaks the condition for destructive interference and generates a measurable light signal at the photodetector.

By monitoring the intensity of this light and its time evolution, the detector records the characteristic imprint of a passing GW. The temporal structure of the measured signal encodes information about the GW amplitude, polarization, and frequency evolution, providing direct evidence for GW propagation and enabling detailed inference of the astrophysical source.

When a GW signal is present in data recorded by an interferometric detector, the measured output inevitably includes contributions from instrumental and environmental noise. The detector output can therefore be modeled as

$$s(t) = n(t) + h(t), \tag{1.33}$$

where s(t) denotes the measured strain time series, $n(t)$ represents the detector noise, and $h(t)$ is the GW induced strain. In the following, we will introduce the basics of signal and detector noises

1.2.1.1 Signal characterization

The response of an interferometric GW detector is quantified by the dimensionless strain $h(t)$, which measures the relative change in the proper lengths of the two detector arms induced by a passing GW. This strain is given by the contraction of the GW metric perturbation $h_{ij}(t)$ with the detector tensor D^{ij},

$$h(t) = D^{ij}h_{ij}(t) = F_+h_+(t) + F_\times h_\times(t), \tag{1.34}$$

where $h_+(t)$ and $h_\times(t)$ are the plus and cross GW polarizations, respectively. The detector tensor is defined as

$$D^{ij} \equiv \frac{1}{2}\left(u^i u^j - v^i v^j\right), \tag{1.35}$$

with u^i and v^i denoting unit vectors along the two interferometer arms.

The functions F_+ and $F_\times$ are the antenna pattern, or beam pattern, functions, which encode the directional and polarization sensitivity of the detector.

We consider an idealized interferometer with two perpendicular arms aligned along the x and y axes of the detector frame (x, y, z). In general, this detector frame does not coincide with the GW frame (x', y', z'), in which the GW is naturally described. Relating the two frames requires a sequence of rotations. First, the GW frame is rotated by an angle ψ about the z' axis to define the polarization basis, aligning the principal axes of the GW polarization ellipse with x' and y'. Next, the frame is rotated by an angle θ about the x' axis and by an angle ϕ about the z' axis, such that the GW propagates toward the detector from the direction specified by the spherical angles (θ, ϕ) in the detector frame.

After applying these rotations, the antenna pattern functions for an interferometer with orthogonal arms are given by

$$F_+^{90^\circ} = \frac{1}{2}(1 + \cos^2\theta)\cos 2\phi \cos 2\psi \cos\theta \sin 2\phi \sin 2\psi, \tag{1.36}$$

and

$$F_\times^{90^\circ} = \frac{1}{2}(1 + \cos^2\theta)\cos 2\phi \sin 2\psi \cos\theta \sin 2\phi \cos 2\psi. \tag{1.37}$$

These expressions show explicitly how the detector response depends on the sky location (θ, ϕ) of the GW source and on the polarization angle ψ. They determine how the GW polarizations project onto a detector, thereby setting the observed strain amplitude and directly influencing detection probability and parameter estimation accuracy. In GW data analysis, antenna pattern functions are therefore essential for modeling selection effects and for coherently combining data from multiple detectors to enable sky localization and polarization reconstruction.

1.2.2 Noise Characterization

For a generic non-Gaussian and non-stationary process, the statistical behavior of the noise cannot be fully characterized by its mean and variance alone, but instead requires knowledge of higher-order moments.

For a real valued stochastic process $n(t)$, which is the detecot noise in this context, the nth moment at times $t_1, \ldots, t_n$ is defined as

$$\mu_n(t_1, \ldots, t_n) \equiv \left\langle \prod_{i=1}^{n} n(t_i) \right\rangle, \tag{1.38}$$

where $\langle \cdot \rangle$ denotes the ensemble average.

Higher-order moments refer to expectations of higher powers of the noise (or products of several noise values at possibly different times), which encode features such as skewness, kurtosis, and more subtle deviations from a Gaussian distribution.

We can also define auto-correlation function

$$R(t_1, t_2) \equiv \langle n(t_1)\, n(t_2) \rangle\,, \tag{1.39}$$

that measures how the detector noise is linearly related to a time-shifted version of itself[1], and it is needed to quantify temporal dependence that the mean and variance alone cannot capture.

In this full generality of noise characterization, the detector noise in GW data will be described with complicated statistical structure, making a rigorous data analysis framework intractable in practice.

To simplify the noise characterization, the detector noise is commonly assumed to be Gaussian, in which case all statistical properties are fully determined by the first and second moments, Since any deterministic offset can be subtracted from the data, the mean can be set to zero without loss of generality

$$\langle n(t) \rangle = 0. \tag{1.40}$$

We can further assume noise to be stationary, then the ensemble average is replaced by time average and the auto-correlation function becomes

$$R(\tau) \equiv \langle n(t + \tau) n(t) \rangle, \tag{1.41}$$

where τ is the time lag of the noise samples.

In ground-based detector, the characteristic correlation time is set by the inverse of the effective noise bandwidth in the sensitive frequency range of the detector. Since ground-based detectors operate primarily over frequencies from a few tens of Hz to a few kHz, the corresponding correlation time is typically of order milliseconds. After standard data conditioning procedures, including bandpass filtering and whitening, most long range correlations are removed, and noise samples separated by $\tau \gtrsim O(10, \mathrm{ms})$ can be treated as effectively uncorrelated, i.e. $R(\tau) \to 0$.

A complementary and more practical description of the noise is obtained in the frequency domain through the one-sided power spectral density $S_n(f)$. The power spectral density is defined as the Fourier transform of the auto-correlation function,

[1] $R(t_1, t_2) = 0$ for uncorrelated noise.

$$\frac{1}{2}S_n(f) \equiv \int_{-\infty}^{\infty} d\tau\, R(\tau)\, e^{i2\pi f\tau}. \tag{1.42}$$

The reality of the detector noise $n(t)$ implies that the product $n(t)n(t+\tau)$ is real, and hence its ensemble average is also real. Consequently, the auto-correlation function $R(\tau)$ is real. This implies $S_n(-f) = S_n^*(f)$

Moreover, stationarity leads to symmetry under time translation, giving $R(-\tau) = R(\tau)$, which implies $S_n(-f) = S_n(f) = S_n(f)$. As a result, the Fourier transform of the auto-correlation function yields a real power spectral density, justifying the definition of the one sided noise spectrum in the frequency domain. Therefore, we have

$$R(0) = \langle n^2(t)\rangle = \int_0^{\infty} df S_n(f). \tag{1.43}$$

This expression shows that the total noise power in the time domain is equal to the integral of the one sided power spectral density over frequency.

An equivalent and frequently used characterization of the noise is obtained directly in the Fourier domain. Denoting the Fourier transform of the noise by $\tilde{n}(f)$, the second order moment is given by

$$\langle \tilde{n}^*(f)\tilde{n}(f')\rangle = \frac{1}{2}\delta(f-f')S_n(f), \tag{1.44}$$

where $\delta(\cdot)$ is the Dirac delta function. This relation makes explicit that different frequency components of the detector noise are statistically uncorrelated and that the power spectral density $S_n(f)$ completely characterizes the noise properties relevant for GW detection and matched-filtering analyses.

Given a one-sided power spectral density, one can define the noise weight inner product

$$\langle a, b\rangle = 4\mathrm{Re}\int_0^{\infty} \frac{\tilde{a}^*(f)\tilde{b}(f)}{S_n(f)} df, \tag{1.45}$$

where a and b denote two arbitrary complex-valued time domain functions, $\tilde{a}(f)$ and $\tilde{b}(f)$ are their Fourier transforms, and $S_n(f)$ is the noise power spectral density of the detector.

This inner product provides a natural metric on the space of signals once the detector noise properties are specified. Frequencies at which the detector noise is small contribute more strongly, while frequencies dominated by noise are automatically down weighted. As a result, the inner product directly quantifies how well a waveform is supported by the detector sensitivity across the observable frequency band.

The noise weighted inner product plays a central role in gravitational wave data analysis. It underlies the definition of the signal-to-noise ratio (SNR), overlaps between waveforms, and mismatch measures used in template bank construction (See details in the next Section). In this sense, detectability of a GW signal is not determined solely by the intrinsic amplitude of the waveform, but by how its spectral content aligns with the frequency dependent sensitivity of the detector as encoded in $S_n(f)$.

1.2.3 Overview of matched-filtering search pipelines

Various GW search pipelines have been developed, each with its own strategy and complementary advantages for identifying signals embedded in noise dominated data These approaches can be broadly classified into modelled searches [31, 32, 33], which rely on accurate waveform predictions, and unmodelled searches [34, 35], which aim to identify generic coherent excess power without assuming a specific signal morphology.

This thesis focuses primarily on modelled searches based on matched filtering, which remain the most sensitive method for detecting compact binary coalescence signals when reliable waveform models are available. Within this framework, the detectability of a GW signal is often approximated through its SNR.

Matched filtering serves as a principal technique for detecting GW signals originating from compact binary coalescences. It represents the optimal linear filtering approach for recovering predefined waveforms immersed in stationary Gaussian noise [22].

In this section, we focus on the matched-filtering search that forms part of the `GstLAL` pipeline [23, 36, 37]. We give particular attention to the signal-consistency test statistic derived from auto-correlation and the ranking statistic applied in `GstLAL`, since these components will play a central role in the chapters presented later. The full matched-filtering search pipeline contains additional elements that are not introduced here. Interested readers are referred to Reference [23] for a detailed description.

1.2.3.1 Template bank

Within a selected region of the search parameter space, for example the component mass plane spanned by m_1 and m_2, a discrete collection of modeled waveforms $h_{T,j}$ is generated in advance. These waveforms, referred to as templates, are then assembled into a template bank. Figure 1.4 shows a template bank utilized by `GstLAL` in the third LVK observing run (O3). Each template, by assuming quasi-circular and aligned spin, is described by a collection of intrinsic parameters $\vec{\theta} = m_1, m_2, \chi_1, \chi_2$, specified in the detector frame [38].

1.2.3.2 Signal-to-noise ratio computation

The matched-filtering output in the time domain, which quantifies the correlation between the detector data d and the template waveform $h_{T,j}$, following Reference [22], is given by

$$z_j(t) = x_j(t) + iy_j(t) = 4\int_0^\infty \frac{\tilde{d}(f)}{\sqrt{S_n(f)}} \frac{\tilde{h}^*_{T,\,j}(f)}{\sqrt{S_n(f)}} e^{2i\pi ft} df, \tag{1.46}$$

where $z_j(t)$ denotes the complex SNR obtained with the jth template, and $x_j(t)$ and $y_j(t)$ are its real and imaginary components, respectively. $\tilde{d}(f)$ and $\tilde{h}_{T,j}(f)$ are the Fourier transforms of the time domain data $d(t)$ and template waveform $h_{T,j}(t)$ respectively, and the asterisk indicates complex conjugation.

The quantity $S_n(f)$ represents the one sided power spectral density of the detector noise. Its square

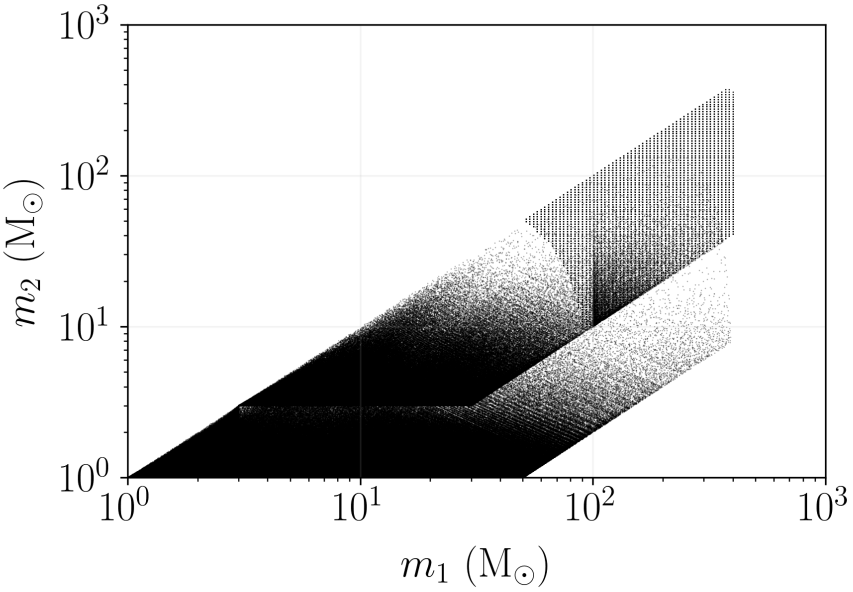

Figure 1.4: Template bank employed in O3 [39]. The bank spans the regions covering binary neutron star, neutron-star-black-hole, and binary black hole systems. In total, it contains approximately 1.8×10^6 templates.

root, $\sqrt{S_n(f)}$, commonly called the noise amplitude spectral density, is applied to whiten both the signal template and the observational data. The factor of 4 appearing in Eq. 1.46 establishes the standard normalization used for the SNR, such that

$$1 = 4 \int_0^\infty \frac{|\tilde{h}_{T,\,j}(f)|}{S_n(f)} df. \tag{1.47}$$

The SNR $\rho(t) = |z_j(t)|$ is defined as the absolute value of the complex output from the matched filter.

To assess how modeling errors decrease the SNR, or equivalently to estimate the proportion of signals that could be undetected in a matched-filtering search, we define the concept of match.

$$\mathcal{M}(h_{T,\,j}, h) = \max_{\phi_0, t_0} \frac{\langle h_{T,\,j}(\phi_0, \tau_0), h\rangle}{\sqrt{\langle h_{T,\,j}, h_{T,\,j}\rangle \langle h, h\rangle}}, \tag{1.48}$$

The phase and time offsets ϕ_0 and τ_0 are selected so as to maximize the match between the template waveform $h_{T,j}$ and the true signal h. Large values of the match $\mathcal{M}$ signify close agreement between $h_{T,j}$ and h, with the upper limit $\mathcal{M} = 1$ attained only when the template reproduces the signal exactly. In the special case $h = h_{T,j}$, and after optimizing the SNR over different noise realizations, the resulting value is called the optimal SNR $\rho_{\rm opt}$, whereas in all other cases the measured value is referred to as the observed SNR $\rho_{\rm obs}$.

The quantity $\mathcal{M}$ in Eq. 1.48 is normalized by the waveform amplitude, while the SNR in Eq. 1.46 depends on both the waveform amplitude and any departures in waveform morphology. Consequently, the match

is sensitive to additional physical effects that alter the form of the waveform but is insensitive to overall rescalings of its amplitude. By contrast, the SNR changes in response to both structural modifications of the waveform and variations in its amplitude.

1.2.3.3 Signal-consistency test

When the SNR $\rho(t)$ rises above a preset threshold $\rho_{\text{threshold}}$, the search pipeline logs the corresponding template parameters as a possible candidate, referred to as a trigger [23]. However, noise transient disturbances in the instrument and departures from Gaussian noise can also produce triggers with large SNR values that imitate true GW events. To mitigate this, `GstLAL` evaluates, for each trigger, a signal-consistency test based statistic χ^2 as an additional supplementary statistics to the SNR.

The signal-consistency test value, ξ^2, quantifies how much the measured SNR time series deviates from the template's expected auto-correlation time series. The auto-correlation time series refers to the SNR time series obtained when the data are filtered with a template that has exactly the same parameters as the signal present in the data.

For a specific template $h_{T,\,j}$, this relationship can be expressed as follows.

$$\xi_j^2 = \frac{\int_{-\delta t}^{\delta t} \left| z_j(t) - z_j(0) R_j(t) \right|^2 dt}{\int_{-\delta t}^{\delta t} 2 - 2 \left| R_j(t) \right|^2 dt}, \tag{1.49}$$

where

$$R_j(t) = \int \frac{|\tilde{h}_{T,\,2j}(f)|^2 + |\tilde{h}_{T,\,2j+1}(f)|^2}{S_n(f)} e^{2\pi i f t} df, \tag{1.50}$$

is the expected SNR time series. In this convention, $t = 0$ marks the instant corresponding to the global maximum of the observed SNR time series. The adjustable parameter δt specifies the extent of the time interval, centered on this peak, over which the integral is evaluated [23]. The denominator is used to scale ξ_j^2 so that a perfect match between the predicted and observed time series gives a value of one, whereas increasing disagreement between them results in larger values of ξ^2.

A useful limiting case is a noise generated trigger that can be modeled as a scaled delta function, $z_j(0)\,\delta(t)$, located at $t = 0$ with amplitude $z_j(0)$. In this scenario, the measured SNR time series is zero everywhere inside the integration interval except at $t = 0$, which produces a pronounced discrepancy relative to the expected time series $z_j(0)\,R_j(t)$. Consequently, the numerator of ξ^2 attains a large value. In addition, raising the SNR of this noise trigger, that is increasing $z_j(0)$, leads to an even larger value of ξ^2. Waveforms that do not accurately follow the template family can also modify ξ^2 in a complicated way, as explored in Reference [40].

1.2.3.4 Likelihood ratio ranking statistics

For each trigger, `GstLAL` computes a ranking statistic given by a likelihood ratio that depends on detection statistics such as ρ and ξ^2 [23, 41, 42, 36, 37]. This ratio quantifies how much more likely a particular combination of these statistics is if a true signal is present rather than only noise. It takes the form

$$\mathcal{L} = \underbrace{\frac{P(\vec{D}_H, \vec{O}, \vec{\rho}, \vec{\xi}^2, \left[\vec{\Delta t}, \vec{\Delta\phi}\right] | \vec{\theta}, \textbf{signal})}{P(\vec{D}_H, \vec{O}, \vec{\rho}, \vec{\xi}^2, \left[\vec{\Delta t}, \vec{\Delta\phi}\right] | \vec{\theta}, \textbf{noise})}}_{(1)} \cdot \underbrace{\frac{P(\vec{\theta}|\textbf{signal})}{P(\vec{\theta}|\textbf{noise})}}_{(2)}. \tag{1.51}$$

In this expression, $\vec{D}_H$ denotes the horizon distances of the individual detectors, and $\vec{O}$ labels which subset of detectors contributed to the trigger. The vector $\vec{\rho}$ contains the measured SNRs in each of these detectors, while $\vec{\xi}^2$ collects the corresponding auto-correlation based signal consistency statistics. The terms $\left[\vec{\Delta t}, \vec{\Delta\phi}\right]$ describe, respectively, the relative arrival time offsets Δt and the differences in coalescence phase $\Delta\phi$ inferred across the detector network.

The first factor (1) expresses the relative probability of obtaining the set of detection statistics for a given trigger under the signal and noise assumptions, while the second factor represents the prior odds on the template parameters $\vec{\theta}$. The $P(\vec{\theta}|\text{signal})$ term in the second factor (2) represents the probability that a trigger produced by a true astrophysical signal is best matched by a template with parameters $\vec{\theta}$. This quantity is commonly referred to as the population model term, as it encodes assumptions about the underlying astrophysical distribution of sources and how these sources map onto the discrete template bank used in the search.

Several earlier works have treated the optimal SNR as a stand in for the detectability of a GW signal can be detected in matched-filtering searches [43, 44, 45]. This assumption, however, does not hold in general. At fixed false-alarm probability, the likelihood ratio is known to be the most powerful choice of detection statistic [46, 42]. The matched-filter SNR can track detectability only when it is approximately proportional to the likelihood ratio, which is true solely under stationary Gaussian noise. Because real detector data depart from this idealized regime, signal detectability must be evaluated through a dedicated injection campaign, rather than by using the matched-filter SNR as a substitute.

`GstLAL` then computes the complementary cumulative distribution

$$C(\ln \mathcal{L}^*|\text{noise}) = \int_{\ln \mathcal{L}^*}^{\infty} P\left(\ln \mathcal{L}|\text{noise}\right) d \ln \mathcal{L}, \tag{1.52}$$

which is then used to evaluate the false-alarm probability, or p-value. This quantity measures the probability that, out of M independent noise like coincidences $\{N_1, N_2, ..., N_M\}$, at least one event attains a log likelihood ratio $\ln \mathcal{L}$ greater than or equal to a specified threshold $\ln \mathcal{L}^*$.

Mathematically, it is written as

$$P(\ln \mathcal{L} \geq \ln \mathcal{L}^* | N_1, ..., N_M) = 1 - \left(e^{-C(\ln \mathcal{L}^* | \text{noise})}\right)^M \tag{1.53}$$

$$= 1 - e^{-MC(\ln \mathcal{L}^* | \text{noise})} \tag{1.54}$$

$e^{-C(\ln \mathcal{L}^* | \text{noise})}$ is the probability that a Poisson process with mean $C(\ln \mathcal{L}^* | \text{noise})$ produces no event with $\mathcal{L} < \ln \mathcal{L}^*$. The false-alarm probability thus captures the chance that noise alone yields a trigger whose likelihood ratio $\mathcal{L}$ is at least as large as that of the candidate, with genuine GW signals expected to give higher likelihood ratios and therefore smaller false-alarm probabilities.

A closely related quantity, the false-alarm rate, is defined for each candidate as

$$\text{FAR} = \frac{C(\ln \mathcal{L}^* | \text{noise})}{T} \tag{1.55}$$

where T is the total analyzed observing time. This rate gives the expected frequency with which noise produces triggers with likelihood ratio $\mathcal{L}$ at or above $\mathcal{L}^*$, so a lower false-alarm rate implies a higher likelihood that the event is astrophysical. In practical `GstLAL` analyses, a commonly used criterion for a confident detection is a false-alarm rate of about 3.85×10^{-7} Hz, corresponding to roughly one spurious event every thirty days.

Chapters 2 and 4 rely heavily on the concepts introduced in this chapter. In particular, Chapter 2 focuses on how new physical effects, such as wave-optics lensing, modify the matched-filter SNR and signal-consistency test statistics. Chapter 4 builds directly on the development of the population model term $P(\vec{\theta}|\text{signal})$ for strongly-lensed GW signals. Key quantities such as the SNR and the false-alarm rate are repeatedly used throughout the thesis.

1.2.4 Current efforts in detectability studies

A variety of methods has been proposed to quantify how detectable GW signals are, each offering different strengths and drawbacks. One widely used approach, employed for example in Refs. [43, 44, 45], computes the optimal network SNR under the assumption that matched-filtering searches recover the true source parameters. This method has also been applied to signals affected by wave-optics effect (See the discussion of lensing in Section 1.4), as in Refs. [40, 47]. While the underlying assumption is not exactly satisfied once noise fluctuations are taken into account, it still serves as a useful practical measure of detectability, particularly when the SNR is large.

Building on this idea, Reference [48] presented a semi analytic formalism for estimating the sensitivity of GW searches. In that study, the distribution of matched-filter SNRs in Gaussian noise is obtained analytically, incorporating correlations across the template bank and maximization over multiple detectors. This technique enables rapid evaluation of detection probabilities without resorting to extensive injection simulations, providing a computationally efficient way to forecast detection rates for compact binary coalescences when dealing with very large populations of potential signals.

In matched-filtering searches, the SNR is only the first step toward identifying convincing GW candidates. Subsequent analysis stages aim to downweight non Gaussian noise fluctuations and construct robust ranking statistics that better separate real signals from instrumental artifacts.

A key element of these pipelines is a signal consistency test, which checks whether the measured data are compatible with the model waveforms in the template bank. If a true GW carries additional physics not encoded in the templates, such as orbital eccentricity, higher order multipoles, beyond general relativity effects, or gravitational lensing signatures, the test can flag it as inconsistent, thereby lowering its effective detection efficiency. A more detailed description of this kind of test, along with concrete examples, has been presented in Section 1.2.3.3.

Since matched-filtering pipelines underpin GW discovery, any mischaracterization of their detection performance can introduce selection effects in population analyses. For instance, reference [49] has examined how signal-consistency tests influence the detectability of beyond general relativity signatures and quantified the resulting selection biases and their consequences for inferred population properties. Studies of this type underscore that search pipelines can preferentially miss or downweight certain classes of signals if their morphology departs from the waveform models used.

The most thorough evaluations of detectability make use of large scale injection campaigns, in which simulated signals are added to detector data and then passed through the full search pipeline. These campaigns have been used to systematically explore how eccentricity [50] and higher order modes [51] affect recovery in matched-filtering searches, providing direct measurements of search efficiency across parameter space. By emulating the entire detection process, they yield detailed information about each injected signal, including recovered SNRs, signal-consistency statistics, and final detection outcomes.

Although such injection studies are computationally intensive, they remain the most reliable means of assessing GW detectability under realistic noise conditions and across a variety of physical scenarios and detector configurations. Their results inform both the interpretation of current observations and the design of future searches, helping to mitigate selection biases introduced by waveform modeling assumptions and signal-consistency tests [23].

1.3 Extracting physics from signals: Bayesian inference

While the previous section focused on matched-filtering searches as a frequentist framework for detecting GW signals and ranking candidate events, detection alone is not sufficient for extracting the full physical content of an observation. Once a signal is identified, a complementary statistical approach is required to infer the source properties and to quantify uncertainties in a principled manner. Bayesian inference provides such a framework by enabling the direct computation of posterior probability distributions for source parameters given the observed data and an explicit noise model.

In this section, we introduce the Bayesian formalism for GW data analysis. We present its application to parameter estimation and hypothesis testing, and we further discuss Monte Carlo methods such as nested sampling. We will introduce simulation-based inference as a novel approach for GW data analysis.

1.3.1 Basics of Bayesian statistics

To extract physical information from GW data, the central objective is to construct the posterior probability distribution over the source parameters,

$$p(\theta|d), \tag{1.56}$$

where d denotes the detector data and θ represents the set of parameters describing a compact binary coalescence. For a generic binary black hole coalescence, θ typically consists of fifteen parameters, including the component masses, spin vectors, sky location, luminosity distance, binary orientation, coalescence time, and phase.

This parameter set should be clearly distinguished from the reduced parameter space employed in matched-filtering search pipelines in Section 1.2.3.1. In matched-filtering searches, waveform templates are often restricted to aligned spin configurations and are characterized only by the component masses and effective spin parameters, resulting in a much lower dimensional space with four intrinsic parameters[2].

The posterior probability distribution is normalized to unity,

$$\int d\theta p(\theta|d) = 1, \tag{1.57}$$

ensuring a consistent probabilistic interpretation. This normalization allows expectation values, credible intervals, and correlations among parameters to be well-defined.

Within this Bayesian framework, parameter uncertainties are quantified using credible intervals derived from the posterior distribution. A credible interval for a parameter or a set of parameters is defined as a region of parameter space that contains a specified fraction of the total posterior probability, such as 90%. For a one dimensional parameter θ, a credible interval $\theta \in [\theta_{\rm low}, \theta_{\rm high}]$ is defined by the condition

$$\int_{\theta_{\rm low}}^{\theta_{\rm high}} p(\theta|d), d\theta = \alpha, \tag{1.58}$$

where α denotes the chosen credibility level to report uncertainties on θ.

1.3.2 Parameter estimation

Bayes theorem states that the posterior probability density of the parameters θ given the data d can be written as

$$p(\theta|d) = \frac{\mathcal{L}(d|\theta)\,\pi(\theta)}{\mathcal{Z}}. \tag{1.59}$$

Here $\mathcal{L}(d|\theta)$ denotes the likelihood function, which quantifies how well a waveform model with parameters θ explains the observed data, and $\pi(\theta)$ is the prior distribution encoding information or assumptions about

[2]Such a reduction is designed to maximize detection efficiency and computational tractability, rather than to provide a complete physical characterization of the source

the parameters before analyzing the data.

The evidence $\mathcal{Z}$ represents the overall probability of the model given the data, marginalized over all model parameters. It is defined as the integral of the likelihood weighted by the prior over the full parameter space,

$$\mathcal{Z} = \int d\theta, \mathcal{L}(d|\theta), \pi(\theta). \tag{1.60}$$

As such, the evidence quantifies how well a given model explains the data after accounting for its allowed parameter volume. Models that fit the data well over a large region of parameter space are favored, while overly flexible models are naturally penalized through this marginalization. The evidence is therefore a key quantity for Bayesian model selection, where ratios of evidences, known as Bayes factors, are used to compare competing physical hypotheses.

Under the assumption that the detector noise is Gaussian and stationary (See Section 1.2.2), the likelihood function admits a simple analytic form

$$\mathcal{L}(d|\theta) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{1}{2}\frac{|d-\mu(\theta)|^2}{\sigma^2}\right]. \tag{1.61}$$

In this expression, d denotes the observed data, $\mu(\theta)$ is the model prediction for a signal with parameters θ, and σ^2 characterizes the noise variance. The exponential term penalizes deviations between the data and the model, weighted by the noise level, while the normalization ensures proper probabilistic interpretation. This Gaussian likelihood forms the foundation of most GW parameter estimation analyses and directly connects the statistical properties of the noise to the inference of source parameters.

1.3.3 Hypothesis testing

In addition to parameter estimation, Bayesian inference provides a natural framework for hypothesis testing and model comparison. This is achieved through the Bayes factor, which quantifies the relative support of the data for two competing hypotheses $\mathcal{H}_1$ and $\mathcal{H}_0$ by taking the ratio of their evidences,

$$\mathcal{B}_{10} = \frac{\mathcal{Z}_1}{\mathcal{Z}_0}, \tag{1.62}$$

where $\mathcal{Z}_i$ denotes the evidence associated with hypothesis $\mathcal{H}_i$.

When prior beliefs about the relative plausibility of the hypotheses are taken into account, the Bayes factor is combined with the prior odds to form the odds ratio,

$$O_{10} = \frac{p(\mathcal{H}_1|d)}{p(\mathcal{H}_0|d)} = \mathcal{B}_{10}\frac{\pi_1}{\pi_0}. \tag{1.63}$$

The odds ratio therefore represents the posterior probability ratio between the two hypotheses given the data, and provides a quantitative criterion for model selection and hypothesis testing.

In Bayesian model comparison, Occam's razor is automatically embodied through the model evidence

(or marginal likelihood). The evidence represents an average of the likelihood over the prior distribution, thereby accounting for both the model's fit to the data and its parameter space volume. Models with more parameters or broader priors are naturally penalized because large portions of their parameter space usually correspond to low likelihood values. Consequently, unless a complex model provides a substantially better fit, simpler models are favored. This mechanism ensures a principled balance between goodness-of-fit and model complexity, preventing overfitting and selecting additional model components only when the data provide strong justification.

1.3.4 Monte Carlo methods

In GW astronomy, Bayesian inference is implemented through a range of sampling techniques designed to explore high dimensional posterior distributions. Widely used software frameworks such as `bilby` provide robust and modular implementations of these methods [52, 53]. These tools have become standard for parameter estimation and model comparison in the LVK analyses [39, 54, 55].

1.3.4.1 Markov Chain Monte Carlo sampling

For Markov Chain Monte Carlo methods, the goal is to generate a sequence of samples whose stationary distribution is the target posterior distribution. This is achieved by constructing a Markov chain that satisfies detailed balance,

$$p(\theta|d)T(\theta \to \theta') = p(\theta'|d)T(\theta' \to \theta), \tag{1.64}$$

where $T(\theta \to \theta')$ denotes the transition probability between states. In practical implementations such as the Metropolis Hastings algorithm, proposed moves are accepted with probability

$$\alpha = \min\left(1, \frac{p(\theta'|d)q(\theta|\theta')}{p(\theta|d)q(\theta'|\theta)}\right), \tag{1.65}$$

where $q(\theta'|\theta)$ is the proposal distribution. These conditions guarantee convergence of the chain to the desired posterior distribution under mild assumptions.

1.3.4.2 Nested sampling

Nested sampling approaches Bayesian inference from a different perspective by focusing on direct evaluation of the evidence. The evidence integral is rewritten in terms of the prior volume X, defined as

$$X(\lambda) = \int_{\mathcal{L}(d|\theta)>\lambda} \pi(\theta)d\theta, \tag{1.66}$$

which measures the fraction of prior mass enclosed above a likelihood threshold λ. With this transformation, the evidence becomes a one dimensional integral,

$$\mathcal{Z} \int_0^1 \mathcal{L}(X) dX. \tag{1.67}$$

By iteratively shrinking the prior volume toward regions of increasing likelihood, nested sampling efficiently estimates the evidence while simultaneously producing posterior samples. These equations highlight the conceptual distinction between MCMC and nested sampling and clarify why both approaches become computationally demanding for high dimensional GW inference problems.

1.3.5 Simulation-based inference

Despite their success, both MCMC and nested sampling methods are computationally expensive for realistic GW analyses. The need to generate large numbers of waveform evaluations in high dimensional parameter spaces leads to substantial runtime and resource demands, particularly for long signals or complex waveform models. These limitations motivate the development of alternative approaches that reduce computational cost, setting the stage for simulation-based inference methods that leverage large scale simulations and amortized learning to accelerate Bayesian inference.

simulation-based inference provides an alternative Bayesian framework that is well suited to problems where likelihood evaluations are computationally expensive or analytically intractable. Rather than evaluating the likelihood explicitly for each parameter point, this approach relies on forward simulations of the data generating process to learn the relationship between model parameters and observables. Once trained, simulation-based inference enables rapid posterior evaluation for new data, making it particularly attractive for high dimensional GW inference problems.

1.3.5.1 Neural posterior estimation

A central class of simulation-based inference techniques is neural posterior estimation. In this approach, a neural network $q_\phi(\boldsymbol{\theta}|d)$ is trained using simulated pairs of parameters and data to directly approximate the posterior distribution $p(\theta|d)$. The network learns a conditional density model that maps an observed dataset to a probability distribution over the source parameters, thereby amortizing the computational cost of Bayesian inference across a large ensemble of simulations rather than performing a costly sampling procedure for each individual observation.

Training is achieved by optimizing the network parameters ϕ such that the learned distribution closely matches the true posterior. This is commonly formulated as the minimization of a divergence between the true and learned posteriors, most often the Kullback-Leibler divergence, leading to the loss function

$$\mathcal{L}(\phi) = \mathbb{E}_{p(\theta,d)}\Big[\mathrm{KL}(p(\theta|d) \parallel q_\phi(\theta|d))\Big], \tag{1.68}$$

where $\mathbb{E}$ denote the expectation.

1.3.5.2 Normalizing flows

Normalizing flows form the most mathematically explicit component of modern simulation-based inference methods and therefore require explicit equations to clearly convey their defining properties.

A normalizing flow is defined through an invertible and differentiable transformation that maps a simple base random variable z to the target variable θ,

$$\theta = f_\phi(z) \qquad z \sim p_0(z). \tag{1.69}$$

Here $p_0(z)$ denotes a tractable base distribution, typically a multivariate Gaussian, and f_ϕ represents a parameterized bijective mapping implemented by a neural network.

The probability density of θ can be evaluated exactly using the change of variables formula,

$$p_\phi(\theta) p_0(z) \left| \det \frac{\partial f_\phi^{-1}}{\partial \theta} \right|. \tag{1.70}$$

This expression highlights the central advantage of normalizing flows: despite their flexibility in modeling complex, non Gaussian distributions, they retain exact and tractable likelihood evaluation through the Jacobian determinant of the inverse transformation. Without these equations, the expressive power and probabilistic consistency of normalizing flows cannot be properly understood.

In contrast, higher level concepts such as amortization, scalability, and reuse across multiple observations can be introduced conceptually without mathematical formalism. In summary, equations are essential for neural posterior estimation and normalizing flows, helpful but optional for introducing simulation-based inference as a general paradigm, and unnecessary for purely computational or motivational discussions

1.3.5.3 DINGO

As a concrete example of simulation-based inference in GW astronomy, we consider `DINGO` (Deep Inference for GW Observations), which employs neural posterior estimation combined with normalizing flows to perform fast and accurate Bayesian inference [56, 57, 58, 59, 60, 61, 62]. `DINGO` is trained on large ensembles of simulated GW signals and directly learns the mapping from detector data to the posterior distribution over source parameters. Once trained, it enables amortized inference without explicit likelihood evaluations, in contrast to traditional sampling based approaches such as those implemented in `bilby`.

Figure 1.5 compares posterior distributions obtained using `DINGO` and a traditional likelihood based method. The green solid contours correspond to posteriors inferred using `DINGO` with a network with 15 source parameter, while the gray dashed contours show results from `bilby` applied to the same injected signal. Note that GW lensing effect is not introduced and we denote any GW signal model without lensing effect as nonlensed. Across the full set of intrinsic and extrinsic parameters, the two methods yield consistent marginal and joint posterior distributions, and both accurately recover the injected parameter values indicated by the vertical reference lines. This agreement demonstrates that simulation-based

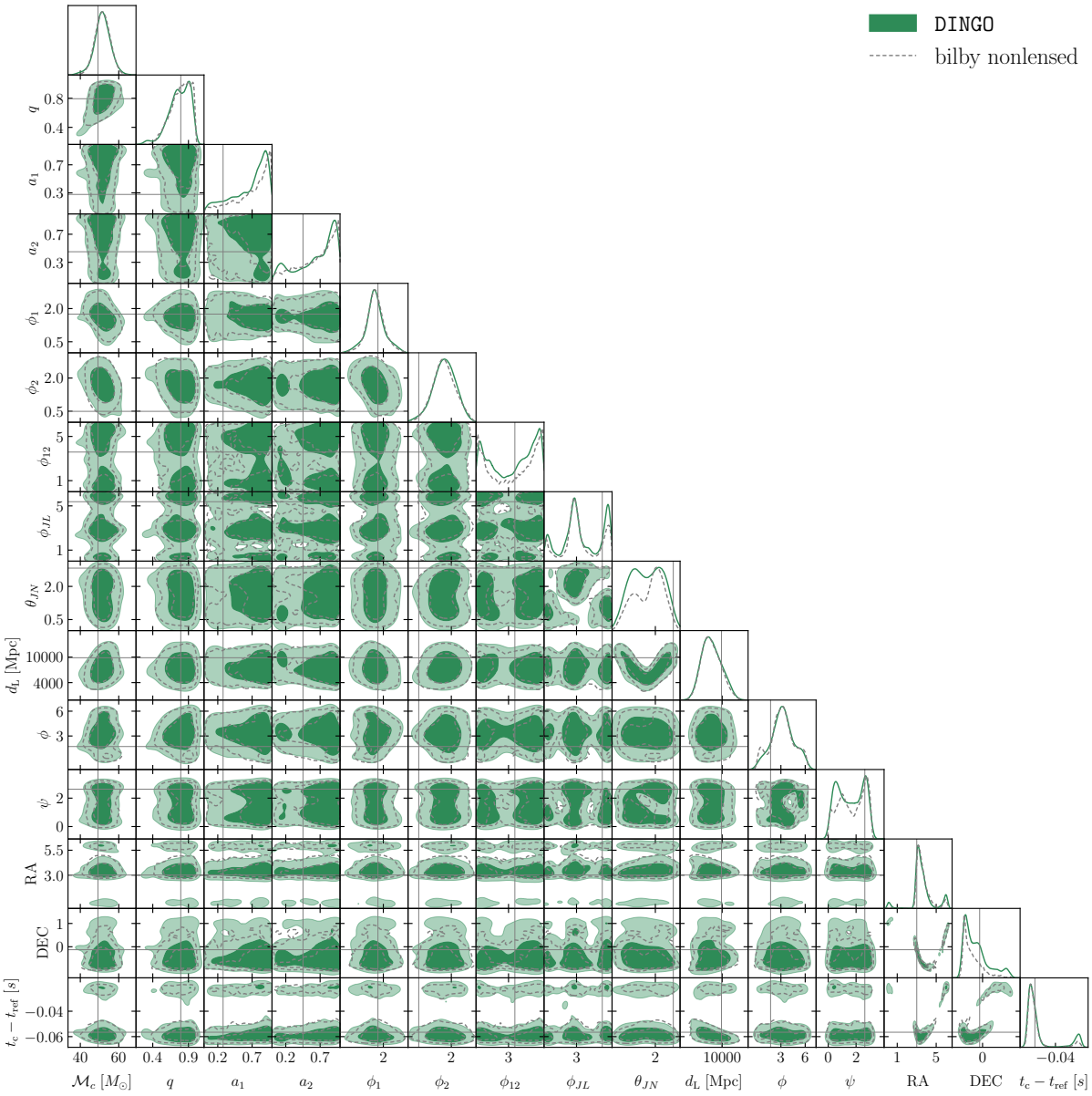

Figure 1.5: Posterior distributions for an injected signal are recovered with `DINGO` using a nonlensed network (green solid curves) and with `bilby` (gray dashed curves). The two inference methods yield consistent posteriors and both accurately recover the injected parameter values, indicated by the solid gray lines and listed in Table 3.1.

inference can reproduce the results of established Monte Carlo methods while dramatically reducing the computational cost at inference time.

The consistency between `DINGO` and traditional Monte Carlo methods highlights the potential of simulation-based approaches for exploring more complex physical scenarios. In particular, new physics effects, such as wave-optics distortions, lead to likelihood surfaces that are expensive or impractical to evaluate with standard methods. By shifting the computational burden to an offline training phase, simulation-based inference provides a scalable framework for incorporating such effects into GW data analysis.

1.4 Gravitational wave lensing

Having established the GW detection method and Bayesian framework for GW data analysis, we now turn to physical effects that can modify GW signals during their propagation [25, 26].

As GWs travel over cosmological distances, they can be bent, magnified, and distorted by massive structures along the line of sight, such as galaxies, galaxy clusters, substrucutre of galaxy clusters or compact objects. Depending on the mass scale and geometry of the lens, gravitational lensing can produce multiple copies of the same signal or induce frequency dependent distortions in a single observed waveform.

In this section, we introduce the theoretical foundations of gravitational lensing as applied to GWs and review current efforts aimed at identifying lensing signatures.

1.4.1 Overview of gravitational lensing

To derive the lensing equation relevant for GWs, we begin by modeling GW propagation as a linear perturbation propagating on top of a smooth background metric. The full metric is decomposed as

$$g_{\mu\nu} = g^{(B)}_{\mu\nu} + h_{\mu\nu}, \tag{1.71}$$

where $h_{\mu\nu}$ represents the GW perturbation and $g^{(B)}_{\mu\nu}$ denotes the background metric sourced by the lens.

Recall the curvature coupled wave equation for GW propagation in arbitrary spacetime in Eq. 1.13:

$$h_{\mu\nu;}{}^{;\alpha}_{\alpha} + 2R^{(B)}_{\alpha\mu\beta\nu}h^{\alpha\beta} = 0, \tag{1.72}$$

where $R^{(B)}_{\alpha\mu\beta\nu}$ is the Riemann tensor associated with the background metric $g^{(B)}_{\mu\nu}$. While the equation remains complicated to solve, we applies several approximations for astrophysical scenario which are commonly adopted.

1.4.1.1 Gravitational lenses in the weak-field regime: Newtonian approximation

In the weak-field regime appropriate for astrophysical lensing, the background metric can be approximated by a perturbed Minkowski spacetime of the form

$$g^{(B)}_{\mu\nu}dx^{\mu}dx^{\nu} = -(1+2U)dt^2 + (1-2U)d\vec{r}^2, \tag{1.73}$$

where $U(\vec{r})$ is the Newtonian gravitational potential generated by the lens mass distribution. For GW lensing, the waves typically propagate far from the central region of compact lenses such as black holes or galactic cores. As a result, the gravitational potential remains small everywhere along the propagation path, satisfying $U \ll 1$.

Under this weak field assumption, the gravitational potential is governed by the Poisson equation

$$\nabla^2 U(\vec{r}) = 4\pi G\rho(\vec{r}), \tag{1.74}$$

where $\rho(\vec{r})$ denotes the mass density of the lens and G is the gravitational constant.

We now assume that the GW wavelength λ is much smaller than the characteristic curvature scale of the background spacetime. In this short wavelength regime, the term proportional to the Riemann tensor represents a higher order correction and can be neglected. The wave equation then simplifies to

$$h_{\mu\nu;\alpha}^{\ \ \ ;\alpha} = 0, \tag{1.75}$$

which is formally identical to the wave equation in a curved but slowly varying background.

1.4.1.2 Gravitational wave as scalar perturbation: Eikonal approximation

When the gravitational field of the lens is weak, the mixing or rotation of the GW polarization states during propagation can be neglected. In this regime, the polarization content is effectively conserved along the ray trajectory. In the short-wavelength, or eikonal / WKB, approximation, the GW perturbation is written as a product of a slowly varying complex scalar amplitude and a polarization tensor,

$$h_{\mu\nu} = \phi e_{\mu\nu}, \tag{1.76}$$

where ϕ encodes both the wave amplitude and phase, and $e_{\mu\nu}$ specifies the polarization state. The polarization tensor $e_{\mu\nu}$ is parallel transported along the null geodesics of the background spacetime, so that its evolution is entirely determined by the background geometry, while all lensing effects relevant for wave optics are captured by the scalar field ϕ.

Substituting the eikonal ansatz in Eq. (1.76) into the linearized wave equation Eq. (1.75), the complex scalar amplitude amplitude satisfy the scalar wave equation on the background spacetime,

$$\partial_\mu \left(\sqrt{-g^{(B)}} g^{(B)_{\mu\nu}} \partial_\nu \phi \right) = 0. \tag{1.77}$$

Here $g^{(B)}$ is the determinant of the background metric $g^{(B)_{\mu\nu}}$. Working in the frequency domain, the scalar field is written as $\phi(\vec{r}, t) = \tilde{\phi}(\vec{r}) e^{-i\omega t}$ where $\omega = 2\pi f$ is the angular frequency. Inserting this into Eq. 1.77 and expanding the metric to leading order in the Newtonian potential U of the lens, one obtains the Helmholtz-type equation

$$\left(\nabla^2 + \omega^2\right) \tilde{\phi} = 4\omega^2 U \tilde{\phi}, \tag{1.78}$$

This equation describes the propagation of GWs in the presence of a weak gravitational potential and provides the starting point for wave-optics treatments of lensing.

To characterize the lensing effect, it is convenient to introduce the amplification factor as the ratio between

the lensed and unlensed GW solutions in the frequency domain

$$F(f) = \frac{\tilde{\phi}^L(f)}{\tilde{\phi}(f)}, \tag{1.79}$$

where $\tilde{\phi}^L$ is the solution of Eq. (1.78) in the presence of the potential U, while $\tilde{\phi}$ denotes the corresponding solution for free propagation with $U = 0$. This dimensionless function $F(f)$ encapsulates all lensing-induced modifications to the GW amplitude and phase as a function of frequency

1.4.1.3 Diffraction intergral: path integral formalism and thin lens approximation

For astrophysical lenses such as galaxies or clusters, the spatial extent of the lens along the line of sight is much smaller than the typical distances from observer to lens and lens to source. Under this thin lens approximation, the gravitational potential U is effectively localized in a plane transverse to the line of sight, and the effect of the lens can be encoded in a two-dimensional projected potential $\Psi(x)$ on the lens plane. The scalar field then accumulates a phase shift when crossing the lens plane, while propagating freely between source–lens and lens–observer in a nearly flat background.

Using the Feynman path integral mehod and angular coordinates $\boldsymbol{\theta}_\mathrm{I}$ and $\boldsymbol{\theta}_\mathrm{S}$ on the image plane (lens plane) and source plane respectively, the GW amplification factor in the frequency domain is obtained as a two-dimensional diffraction integral

$$F(f, \boldsymbol{\theta}_\mathrm{S}) = \frac{\tau_\mathrm{D} f}{i} \int \mathrm{d}^2\boldsymbol{\theta}_\mathrm{I} \, \exp[2\pi i f \, t_\mathrm{d}(\boldsymbol{\theta}_\mathrm{I}, \boldsymbol{\theta}_\mathrm{S})] \,, \tag{1.80}$$

where the integration is carried out over the image plane. The vector $\boldsymbol{\theta}_\mathrm{S}$ denotes the nonlensed angular position of the source on the sky. The phase of the integrand is determined by the relative time delay t_d between different paths connecting the source and the observer, often referred to as the time delay surface. This formulation explicitly accounts for the coherent superposition of all possible propagation paths, which is essential for capturing wave-optics effects.

$$t_\mathrm{d} \approx \tau_\mathrm{D} \left[\frac{1}{2} |\boldsymbol{\theta} - \boldsymbol{\theta}_\mathrm{S}|^2 - \Psi(\boldsymbol{\theta}) \right] \tag{1.81}$$

where each term corresponds to the geometric and Shapiro delays respectively, with $\Psi = 2/\tau_\mathrm{D} c^3 \int \phi ds$ and c is the speed of light. and ϕ is the Newtonian potential of the lens.

We define τ_D as the characteristic timescale associated to the distances between source-lens-observer

$$\tau_\mathrm{D} = \frac{D_L D_S}{c D_{LS}}, \tag{1.82}$$

where D_L, D_S, and D_{LS} denote the angular diameter distances from observer to lens, observer to source, and lens to source respectively.

To further simplify the expression, we rewrite the angular frequency and time delay by introducing a reference angular scale θ_*

$$w \equiv \tau_D \theta_*^2 \omega, \tag{1.83}$$

$$T_d \equiv \frac{t_d}{\tau_D \theta_*^2}. \tag{1.84}$$

We can now define dimensionless source and image position using $\boldsymbol{y} = \boldsymbol{\theta_S}/\theta_*$ and $\boldsymbol{x} = \boldsymbol{\theta_I}/\theta_*$.

The dimensionaless time delay becomes

$$T_d(\boldsymbol{x}, \boldsymbol{y}) = \phi(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{2}|\boldsymbol{x} - \boldsymbol{y}|^2 - \psi(\boldsymbol{x}), \tag{1.85}$$

where the $\psi = \Psi/\theta_*^2$ is the dimentionless lens potential that satisfy the two-dimensional Poisson equation

$$\nabla_{\boldsymbol{x}}^2 \psi(\boldsymbol{x}) = \frac{2\Sigma}{\Sigma_{\rm cr}}. \tag{1.86}$$

Here Σ is the surface mass density of the lens and $\Sigma_{\rm cr} = D_S/(4\pi D_L D_{LS})$.

The diffraction integral can be cast into a dimensionless form as

$$F(w, \boldsymbol{y}) = \frac{w}{2\pi i} \int \mathrm{d}^2\boldsymbol{x}\ \exp[iwT_d(\boldsymbol{x}, \boldsymbol{y})]\,. \tag{1.87}$$

This formulation makes explicit that the amplification factor depends only on the dimensionless frequency parameter w and the scaled source position $\boldsymbol{y}$, independent of the overall physical length scales.

For a given surface mass density Σ, the lensing potential and hence the time delay function can be obtained by solving the corresponding Poisson equation and substituting the result into the diffraction integral. In practice, however, closed form expressions for $F(w, \boldsymbol{y})$ exist only for a limited set of highly idealized mass distributions, such as point mass or singular isothermal profiles. For more realistic lenses, the integral cannot be evaluated analytically and must instead be treated using approximation schemes.

In certain limit, the diffraction integral can be systematically approximated using different approximations. A commonly adopted stationary phase approximation is introduced in the next section.

1.4.1.4 Geometrical optics: Stationary phase approximation

When the phase factor $\exp[iwT_d(\boldsymbol{x}, \boldsymbol{y})]$ becomes highly oscillatory, namely in the regime $wT_d \gg 1$, contributions from most regions of the integration domain cancel out due to rapid phase variations. As a result, the integral is dominated by the stationary points $\boldsymbol{x}_j$ of the time delay surface, where the gradient of $T_d(\boldsymbol{x}, \boldsymbol{y})$ vanishes and the phase varies least rapidly.

$$\left.\frac{\partial T_d}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}_j} = 0. \tag{1.88}$$

We can Taylar expand $T_d$round $\boldsymbol{x} = \boldsymbol{x}_j$ and keep upto the quadratic term

$$T_d(\boldsymbol{x}) = T_d(\boldsymbol{x}_j) + \frac{1}{2}\frac{\partial^2 T_d}{\partial x_a \partial x_b}\Big(x_a - x_{j,a}\Big)\Big(x_b - x_{j,b}\Big), \tag{1.89}$$

the diffraction intergral become a two dimensional Gaussian integral which we can solve easily and obtain the amplication factor as

$$F(w) \approx \sum_j \sqrt{|\mu(\boldsymbol{x}_j)|}\exp\Big(iwT_d(\boldsymbol{x}_j) - in_j\pi/2\Big). \tag{1.90}$$

Hence, different stationary point represent distinct images arrive at different time T_j with magnification μ_j, and we referred it as geometrical optics limit. The phase shift n_j is also known as the Morse phase facor. Define the Hessian matrix of the time delay surface

$$T_{ab} \equiv \frac{\partial^2 T_d}{\partial x_a \partial x_b}, \tag{1.91}$$

the lensing magnification is

$$\mu_j = 1/\det(T_{ab}(\boldsymbol{x}_j)) \tag{1.92}$$

Each lensed image corresponds to a stationary point of the time delay function. At that point, the Hessian matrix has two eigenvectors with associated eigenvalues that determine the local curvature of the time delay surface. The signs of these eigenvalues classify the image as a minimum, saddle, or maximum:

$$n = \begin{cases} 0, \text{ for local minimum (two positive eigenvalues, Type 1)} \\ 1, \text{ for saddle point (one positive and one negative eigenvalue, Type 2)} \\ 2, \text{ for local minimum (two negative eigenvalues, Type 3)} \end{cases}, \tag{1.93}$$

When $\det(T_{ab}) \to 0$, corresponding to the degeneracy of image positions, as the associated magnification diverges ($\mu_j \to \infty$), thereby signaling the breakdown of the stationary phase treatment. In this regime, the Hessian matrix of the time delay function becomes singular and the quadratic expansion around a stationary point is no longer sufficient to approximate the phase. The critical curves are therefore defined by the condition

$$\det(T_{ab}) = 0, \tag{1.94}$$

which allows one to identify the locations in the image plane where this degeneracy occurs and to infer the associated caustic structure in the source plane.

Next, we will illustrate two examples of lens models, point-mass lens and overlapping chirps with stationary phase approximation, which are utlized in the following chapters.

1.4.1.5 Point-mass lens

The simplest lens model one can get is the point-mass lens, which correspond to the surface mass density $\Sigma(\boldsymbol{x}) = M_L\delta^2(\boldsymbol{x})$ and lens potential $\psi(\boldsymbol{x}) = \ln x$. The reference angular scale θ_* is charaterized by the lens mass M_L, the reference scale is the Einstein angle

$$\theta_E = \sqrt{\frac{4GM_L}{c^2}\frac{(1+z_L)D_{LS}}{D_L D_S}}\,. \tag{1.95}$$

Hence, the dimensionless source position is simplified under spherical symmetry as, and depends on the impact parameter $y = \theta_S/\theta_E$. Here we further define the redshifted lens mass in the constant as $M_{Lz} = M_L(1+z)$ The amplification factor can be solved analytically and given as

$$F(f) = \exp\{\frac{\pi w}{4} + \mathrm{i}\frac{w}{2}[\ln(\frac{w}{2} - 2\phi_m)]\}\Gamma(1 - \frac{\mathrm{i}w}{2})_1F_1(\frac{\mathrm{i}w}{2}, 1; \frac{1wy^2}{2}), \tag{1.96}$$

where $\phi_m = (x_m - y)^2/2 \; - \ln x_m$, $x_m = (y + \sqrt{y^2+4}/2)$ and ${}_1F_1$ is the confluent hypergeometric function. The point-mass lens in this closed form is utlized in Chapter 2.

Using stationary phase approximation, the amplification factor becomes

$$F(f) = |\mu_+|^{1/2} - \mathrm{i}|\mu_-|^{1/2}e^{2\pi\mathrm{i}f\Delta t_d}, \tag{1.97}$$

where the magnification and time delay for each image are

$$\mu_\pm = \frac{1}{2} \pm \frac{y^2+2}{2y\sqrt{y^2+4}}\,, \tag{1.98}$$

$$\Delta t = \frac{4GM_{Lz}}{c^3}\Delta T(y), \tag{1.99}$$

where $T(y)$ is the dimensionless time delay

$$\Delta T(y) = \frac{1}{2}y\sqrt{y^2+4} - \ln\left|\frac{\sqrt{y^2+4}+y}{\sqrt{y^2+4}-y}\right|. \tag{1.100}$$

In this geometrical optics limit of point-mass lens, Type 1 and type 2 are always formed with a $\pi/2$ phase difference. The stationary phase approximation of point-mass lens is adopted in Chapter 3

1.4.1.6 Two-image inteference

Consider a lensing configuration that produces two images with a relative phase shift of $\pi/2$, a relative magnification $\mu_{\rm rel}$, and a time delay Δt. The resulting amplification factor can be written as

$$F(f>0) = 1 + \sqrt{\mu_{\rm rel}}\,e^{i(2\pi f\Delta t - \pi/2)}\,. \tag{1.101}$$

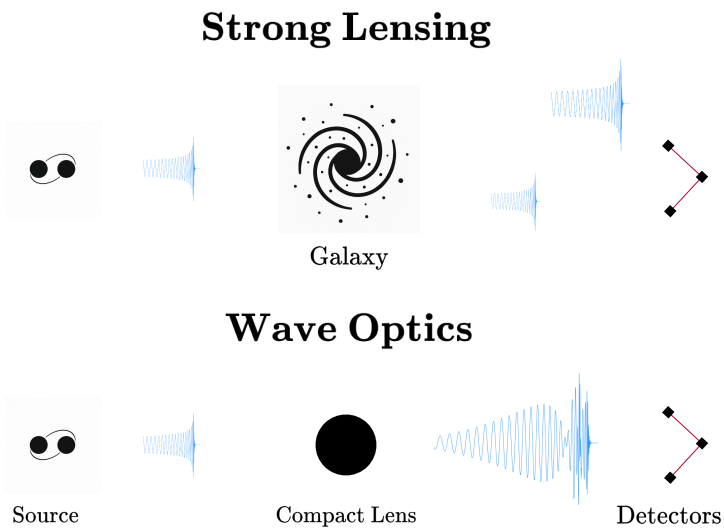

Figure 1.6: Illustration of two gravitational lensing regimes relevant for GWs. In the strong lensing regime (top), a massive galaxy deflects the signal and produces multiple macroscopic images that arrive at the detectors as repeated, magnified copies of the original waveform. In the wave-optics regime (bottom), a compact lens with a much smaller mass does not form resolvable images but instead modulates the waveform through diffraction and interference, imprinting characteristic frequency dependent distortions. Together, these regimes highlight how GWs encode lensing effects across a wide range of astrophysical lenses.

This expression captures the interference between the two dominant images and is formally equivalent to the description commonly used in the strong lensing regime.

The form in Eq. (1.101) provides a reliable approximation whenever the observed waveform is governed by the coherent superposition of two images. It has been shown to remain valid across a broad class of lensing singularities, including point, fold, and certain cusp configurations, where the local lensing geometry effectively reduces to a two image interference problem. Detailed discussions of its domain of applicability and limitations can be found in Refs [63, 64].

Given its simplicity and clear physical interpretation, this parametrization is adopted in Chapter 3 as a convenient and flexible model for capturing wave-optics effects.

1.4.2 Phenomenology of gravitational waves lensing

In the previous section, the theoretical framework was presented in a form that is equally applicable to electromagnetic waves and GWs. However, differences in wavelength and detector characteristics imply that lensed GWs can be observed in lensing regimes (See Fig. 1.6) that are distinct from those typically accessible in electromagnetic observations. In particular, for ground-based GW detectors, the characteristic wavelength of the signal is of order 10^6 m, which is many orders of magnitude larger than that of electromagnetic radiation. This long wavelength makes GWs especially sensitive to diffraction and interference effects, allowing access to lensing regimes that are effectively unobservable with electromagnetic waves.

In this section, we introduce the different phenomenologies of GW lensing and discuss their applications to cosmology and tests of general relativity.

1.4.2.1 Wave-optics lensing

When the characteristic size of the lens becomes comparable to the GW wavelength, $R_{\rm Lens} \sim \lambda_{\rm GW}$, the wave nature of the signal can no longer be neglected. In this regime, lenses such as intermediate-mass black holes or dark matter substructures embedded in galaxies can imprint frequency-dependent modulations on the observed waveform. These effects originate from diffraction and interference between multiple propagation paths and manifest as lensing induced distortions within a single GW signal rather than as clearly separated multiple images [65, 66, 26, 67, 68]. Such wave-optics features have been extensively studied and provide a direct probe of small scale gravitational potentials along the line of sight.

The presence of these modulations opens up a range of scientific applications. The detailed structure of the distortions encodes information about the abundance and mass spectrum of compact lenses [69, 70, 71, 72, 73, 74, 75, 76, 77, 78], the caustic strucutre associated with different lens models [13], the underlying halo mass function [79, 80, 81, 82] and the strong-field regime of gravity [6, 83, 84, 85, 86, 87]. On cosmological scales, the frequency-dependent phase and amplitude variations also carry sensitivity to the cosmic expansion history through their dependence on lens and source distances [88, 89, 90]. From a data analysis perspective, accurately modeling these effects is essential, since unmodeled lensing distortions can bias parameter estimation [91, 68] and potentially mimic or obscure genuine signatures of deviations from general relativity [92, 93, 68].

An illustrative example is shown in Fig. 1.7, which presents a GW lensed by a point-mass. In the time domain, the lensed signal exhibits characteristic beating patterns and amplitude modulations relative to the unlensed waveform, reflecting the interference between multiple paths with different time delays. In the frequency domain, these effects appear as a series of oscillatory features superimposed on the smooth unlensed spectrum, with the modulation scale set by the lens mass and the source lens geometry. This example highlights how wave-optics lensing produces correlated signatures in both time and frequency domain, providing a clear observational handle on lensing by compact objects in the regime $R_{\rm Lens} \sim \lambda_{\rm GW}$.

1.4.2.2 Strong lensing

When the source, lens, and observer are nearly collinear and the characteristic lens scale satisfies $R_{\rm Lens} \gg \lambda_{\rm GW}$, GWs propagate along well defined rays and undergo achromatic magnification. This situation is relevant for galaxy scale lenses with $R_{\rm Lens} \sim 10^{20}$, m and for galaxy clusters with $R_{\rm Lens} \sim 10^{23}$, m, which act as the dominant lens populations in this regime [94, 95, 96, 14]. The observed GW signal then consists of multiple copies of the same intrinsic chirp, separated in arrival time and differing only by constant phase shifts, relative magnifications, and relative time delays (See Section 1.4.1.4). In strong lensing regime of GW, multiple images are produced and expected to be detected as separate GW events [97, 98].

strongly-lensed GWs enable a broad range of applications, including precision tests of general relativity through consistency checks across images [99, 100, 101, 102], improved sky localization from repeated

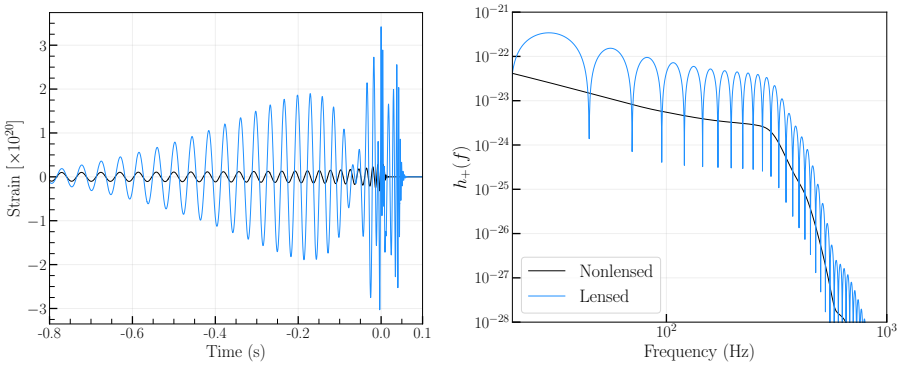

Figure 1.7: Example of wave-optics GW lensing by a point mass with impact parameter $y = 0.01$ and redshifted lens mass $M_{\mathrm{Lz}} = 10^5\ M_\odot$. The left panel shows the time domain strain, where interference between multiple propagation paths produces characteristic amplitude modulations relative to the nonlensed signal. The right panel displays the corresponding frequency domain strain, illustrating oscillatory features imprinted on the spectrum by diffraction and interference effects when the lens scale is comparable to the GW wavelength.

detections [103, 104], early warning of subsequent images once a first signal is observed [105], and cosmological inference using time delay measurements [106, 107, 108, 109], including constraints on the dark energy equation of state [110] and lensing cosmography [111, 112, 113, 114]. In addition, strongly-lensed events can appear as population outliers in GW catalogs [115] and provide constraints on the properties of the lensing galaxy population itself [116, 9, 117].

Current rate estimates based on galaxy scale lenses suggest that strong lensing occurs at the level of roughly one event per 10^3 detected GWs, although this remains sensitive to assumptions about source populations, lens distributions, and detector sensitivity [118, 119, 120, 98, 121, 122, 123, 124].

1.4.3 Current efforts in searches for lensed gravitational waves

Since the first direct detection of GWs by the LVK Collaboration in 2015 [125, 126], extensive efforts have been devoted to searching for lensing signatures within the detected event population reported in the GWTC. Up to GWTC 3.0, these searches have not yielded statistically compelling evidence for GW lensing [127, 20, 16, 17, 12, 5]. Multiple analyses have systematically investigated both strong lensing and wave-optics effects, including repeated events, interference patterns, and magnification outliers, in catalogs of compact binary coalescences [128, 18, 129]. One candidate reported in Reference [130] was later deemed astrophysically implausible. Consequently, existing studies have primarily reported upper limits or null results on lensing rates rather than confirmed detections. In GWTC 4.0, the event GW231123_135430, hereafter GW231123, is identified as the highest ranked single event lensing candidate in current LVK searches[55, 131, 5]. This event has motivated several dedicated follow up analyses [3, 132, 133, 134],

which are discussed in detail in Section 3.3.2.

While a broad range of lensing analyses and methodologies has been established, significant challenges remain in the detection of lensed GW signals [127, 20, 16, 17, 12, 5]. Existing studies are restricted to events that have already been identified by standard detection pipelines optimized for nonlensed compact binary coalescence signals. These pipelines employ template banks constructed under the assumption of unlensed waveforms and are designed to recover the widest possible generic compact binary population. When lensing induces strong waveform distortions, particularly in the wave-optics regime, the overlap between lensed signals and unlensed templates can be substantially reduced, causing some lensed events to fall below standard detection thresholds. If such signals are systematically missed at the detection stage, lensing analyses that operate only on detected events become biased. In this case, inferred lensing rates or constraints derived from null results may underestimate the true occurrence of lensing, since a fraction of the lensed population never enters the observed catalog. The detectability of lensed GWs depends jointly on source parameters such as masses, spins, and distance, as well as lens parameters including mass, impact parameter, redshift, and lens geometry. Because current template banks do not model lensing effects, it is necessary to explicitly quantify how lensing modifies matched filter SNR, signal consistency statistics, and false-alarm rates across this combined parameter space.

A robust claim of GW lensing also requires the assessment of statistical significance. A commonly adopted approach relies on Bayesian model selection between lensed and unlensed hypotheses. While the resulting lensing Bayes factor quantifies the relative support for the lensing hypothesis, it does not by itself establish the significance of a detection, since waveform systematics and noise fluctuations can also produce apparent support for lensing. Properly accounting for these background contributions requires large ensembles of analyses to estimate fals-alarm probability and significance levels such as 3 or 5 σ. However, performing thousands or millions of Bayesian inferences using traditional Monte Carlo methods is computationally prohibitive due to their high cost [52, 135, 136]. As a consequence, up to GWTC 3.0, most lensing investigations have not reported statistically robust significances or lensing false-alarm probability, leaving the discovery of lensed GWs without a clear statistical benchmark [127, 20, 16, 17, 12, 5].

An additional challenge arises in searches for strong lensing based on multiple images. As the number of detected GW events increases, the number of possible event pairs grows rapidly [137, 138]. This leads to an increasing probability that unrelated events with similar source properties are misidentified as lensed pairs. To mitigate this issue, several studies have developed alternative statistical frameworks for identifying multiple images [139, 140, 141], incorporated lens catalogs from electromagnetic surveys [14], or exploited waveform distortions from overlapping images as distinctive evidence of lensing. Since some lensed images can be significantly demagnified, extending searches below standard detection thresholds can further reduce false associations and improve confidence in strong lensing identification.

This thesis is organized to address these challenges in a systematic manner. The impact of lensing induced waveform distortions on GW detectability in matched-filtering seaches is studied in Chapter 2. The problem of performing rapid and accurate inference for lenisng is tackled using simulation-based inference techniques, which are addressed in Chapter 3. Finally, a new method of searching for strongly-lensed sub-threshold images is detailed in Chapter 4.

Chapter 2

Detectability of lensed GWs in matched-filtering searches

The first step in understanding how new physics imprints observational signatures in the GW detectors is to evaluate the mismatch against waveforms without the physics and the SNR. In particular, the SNR plays a central role in matched-filtering searches, quantifying how loudly a GW stands out against detector noise. It is, therefore, widely used as a proxy for the detectability of exotic or beyond standard signals. This forms the basis of rate estimates and population studies across GW astrophysics [142].

In this chapter, we use gravitational lensing as an example of how a signal can be distorted and magnified. We present the first investigation of the detectability of lensed GWs from compact binary coalescences using a matched-filtering search pipeline, `GstLAL`. We employ the point-mass lens model throughout the study, (See introduction to point-mass lens model in Section 1.4.1.5) which is specified by impact parater y and the redshifted lens mass M_{Lz}. The lensing amplicaition factor is given in Eq.1.96. Contrary to expectations based on optimal SNR arguments, we show that a louder signal does not necessarily lead to higher detectability when a template bank does not include the lensing effects. Lensed signals in the wave optics regime can be heavily penalised by signal consistency tests, reducing their detection efficiency from about 90% in the nonlensed case to $< 1\%$ when lensing induced distortions are present.

These results highlight the need to extend current template banks to capture lensing effects and to revisit existing constraints on compact dark matter. They also demonstrate that without such improvements, lensed GWs may already be passing through our detectors unnoticed.

This chapter is organized as follows. Section 2.1 reviews current efforts to assess the detectability of GWs. Section 2.2 outlines the design of our injection campaigns, which examine (1) how gravitational lensing affects matched-filtering searches and (2) how these effects correlate with signal amplitude. Together, these studies allow us to predict the general detectability of lensed GWs. Section 2.3 presents the results from each campaign. Finally, Section 2.4 summarises the main findings and discusses their implications for ongoing lensing analyses.

2.1 Current efforts of detectability of lensed GWs: Match and optimal SNR

To quantify how lensing within the chosen parameter space alters the waveform morphology, both the match between the lensed and nonlensed signals, $\mathcal{M}(h_{\rm L}, h_{\rm NL})$, and the ratio of optimal SNRs, $\rho_{\rm L,opt}/\rho_{\rm NL,opt}$, are evaluated. Here $\rho_{\rm L,opt}$ is the optimal SNR obtained from the lensed waveform, while $\rho_{\rm NL,opt}$ denotes the corresponding optimal SNR for the nonlensed waveform. In both cases, a nonspinning 30–30 $M_\odot$ waveform is used as the fiducial signal. As explained in Sec. 1.2.3, the match $\mathcal{M}(h_{\rm L}, h_{\rm NL})$ is normalized by the amplitudes of both waveforms and therefore does not respond to lensing magnification, whereas $\rho_{\rm L,opt}$ reflects the combined impact of waveform distortion and amplitude amplification. Consequently, whenever magnification dominates over morphological changes, one expects $\rho_{\rm L,opt}/\rho_{\rm NL,opt} > 1$.

The left panel of Fig.2.1 demonstrates that the match, using SEOBNRV4_ROM approximant, becomes particularly small in regions with large lens mass $M_{\rm Lz}$ and small impact parameter y. In this part of parameter space, the associated lensed signals develop strong beating patterns caused by large magnifications and short time delays, as illustrated in the right panel of Fig. 2.2. In such regimes, attempting to analyze lensed signals with nonlensed templates leads to substantial mismatches.

By construction, the optimal SNR assumes that the template achieving the highest match is identical to the true signal, and higher optimal SNR values are typically associated with better detectability. For point-mass lenses in the wave-optics regime, magnification is always present and thus enhances the optimal SNR. This behavior is visible in the right panel of Fig. 2.1, where the optimal SNR grows with the strength of the lensing, i.e. as a function of the lens parameters. In particular, combinations of large lens mass $M_{\rm Lz}$ and small y, which produce stronger magnification, yield larger optimal SNRs for the lensed signals. The right panel of Fig. 2.2 clarifies the different trends in match and SNR seen in Fig. 2.1: a single point-mass lens can boost the waveform amplitude by a factor of at least five while still giving $\mathcal{M}(h_{\rm L}, h_{\rm NL}) \simeq 0.7$, corresponding to a situation where magnification dominates over distortion and $\rho_{\rm L,opt}/\rho_{\rm NL,opt} \sim 10$. This qualitative picture agrees with the matched-filtering study of Reference [47].

In practical matched-filtering searches, signal detectability is often approximated using observed SNRs, which are frequently inferred from optimal SNR values. In realistic observing conditions, however, detector noise is time dependent, and the template that yields the largest SNR for a given signal, denoted $h_{\rm tem}$, can differ from the true waveform $h_{\rm true}$. As a consequence, both the observed SNR and the optimal SNR can deviate from their noise free expectations. Beyond noise fluctuations, lensing induced distortions of the waveform further complicate the search pipeline's ability to identify the best matching template, leading to nontrivial behavior of both observed and optimal SNR. For these reasons, injection campaigns are essential to robustly determine how gravitational lensing affects signal detectability in matched-filtering searches.

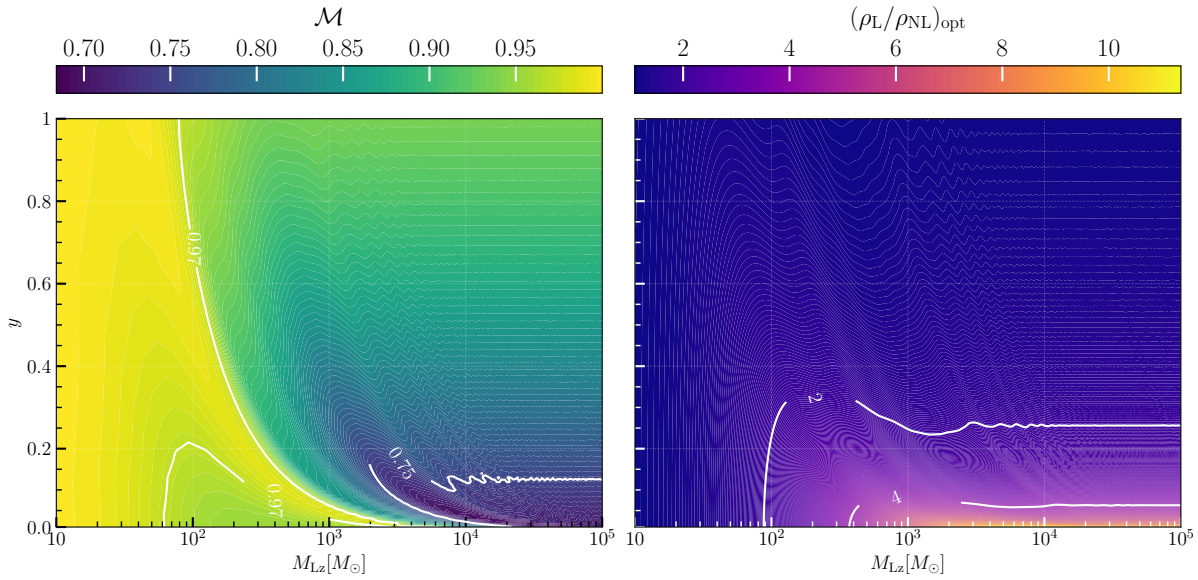

Figure 2.1: Left panel shows the match between lensed signals and nonlensed templates, $\mathcal{M}(h_L, h_{NL})$, as a function of the redshifted lens mass M_{Lz} and the impact parameter y. Right panel shows the ratio of the optimal SNR for lensed and unlensed signals, $\rho_{L,opt}/\rho_{NL,opt}$, over the same (M_{Lz}, y) parameter space. The calculation uses a nonspinning binary black hole source at a luminosity distance of 125 Mpc with component masses $m_1 = m_2 = 30\,M_\odot$. For the point-mass lens model, the condition $\rho_{L,opt}/\rho_{UL,opt} > 1$ is always satisfied.

2.2 Methods

Three injection campaigns are performed with the time domain matched-filtering search pipeline `GstLAL` to investigate how gravitational lensing influences the detectability of GW signals. The resulting distributions of recovered source parameters, matched-filter SNR, signal-consistency statistics, and false-alarm rates reveal how waveform modifications and stochastic noise jointly shape search performance as the source parameters are varied. The specific search configuration and details of the injected signal used in these studies are summarized in Table 2.1.

2.2.1 Matched-filtering searches setup

In matched-filtering searches, evaluating the likelihood ratio for injected signals accurately requires a sufficiently large sample of background noise events to estimate the denominator in Eq. 1.51. Noise fluctuations can also cause an injection to be recovered by a template whose source parameters differ from those of the true signal, which in turn necessitates that the template bank cover a parameter region larger than the injection set itself.

To focus on lensing effects, this study considers a representative nonspinning binary black hole system with component masses of thirty solar masses each. Under this choice, the main independent variables are the lensing parameters, which determine the strength of lensing, and the source distance, which fixes the intrinsic signal strength, allowing construction of a relatively compact yet efficient template bank

Set	Type	Campaign	y	M_{Lz} [$M_\odot$]	Distance [Mpc]	Found
1	Nonlensed	1,2,3	...	...	125	17020
2	Lensed	1	0.01	10	125	17068
3	Lensed	1	0.01	10^3	125	14251
4	Lensed	1,2	0.01	10^5	125	1
5	Nonlensed	2,3	...	...	625	12878
6	Nonlensed	2,3	...	...	1250	6217
7	Lensed	2	0.01	10^5	625	8
8	Lensed	2	0.01	10^5	1250	13
9	Lensed	3	[0.01,1]	[10,10^5]	125	14303
10	Lensed	3	[0.01,1]	[10,10^5]	625	12536
11	Lensed	3	[0.01,1]	[10,10^5]	1250	8254

Table 2.1: Set of 11 injection configurations employed in the three simulation campaigns. Each set contains 19,432 signals. Campaign 1 quantifies lensing effects on detectability, Campaign 2 isolates signal strength dependence, and Campaign 3 maps overall detection performance. An injection is classified as found if its false-alarm rate is below 3.85×10^{-7} Hz. All waveforms correspond to a nonspinning 30–30 $M_\odot$ binary black hole source.

following earlier strategies [143].

The template bank spans binary black hole component masses with m_1 between ten and ninety solar masses and m_2 between ten and forty solar masses, and contains a total of fourteen thousand six hundred six templates, ensuring that all injections lie comfortably within the covered parameter space. Consistent with GWTC-3, the SEOBNRV4_ROM approximant is used for the templates [144], including only the quadrupole mode and aligned spins, with spin components drawn uniformly from −0.99 to +0.99 [39].

Using a reduced template bank introduces a potential limitation in estimating the false-alarm rate in a detectability study, since the background depends on how well source parameters are recovered. Increasing the bank size mainly enlarges the region in which parameter recovery can be inaccurate while greatly increasing computational cost, so a smaller bank is adopted as a pragmatic compromise.

The analysis uses data from the LIGO Hanford and Livingston detectors during the first week of O3, with data quality selection and glitch mitigation aligned with the GWTC-3 procedures [39]. To assess the impact of lensing on matched-filtering searches, the distributions of recovered chirp mass, effective spin, matched-filter SNR, the `GstLAL` autocorrelation-based signal-consistency statistic ξ^2, and the resulting false-alarm rate are examined, with a representative detection threshold corresponding to a false-alarm rate of 1 in 30 days.

2.2.2 Injection campaigns

These injection campaigns start with two runs specifically constructed to disentangle the roles of lensing and intrinsic signal strength in determining detectability. The resulting baseline characterizes how each effect operates in isolation and guides the interpretation of the more realistic third campaign. A total of eleven injection sets is generated, each comprising 19, 432 simulated signals produced with the SEOBNRV4 waveform model [144] implemented in `LALSimulation` [145].

For every injection, the sky location (right ascension and declination), inclination, and polarization are drawn from uniform distributions, and the injections are placed uniformly in time over the analyzed data

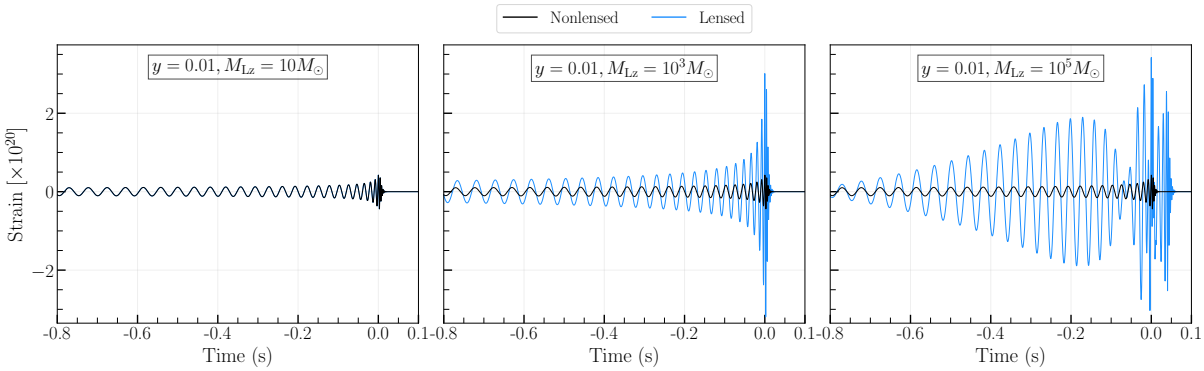

Figure 2.2: The time-domain GW signals from a nonspinning binary black hole with component masses $m_1 = m_2 = 30\,M_\odot$ are shown in black for the nonlensed case and in blue for the lensed case, demonstrating increasingly strong amplification. All lensed waveforms are computed with a common impact parameter $y = 0.01$. The left panel corresponds to a redshifted lens mass $M_{\rm Lz} = 10\,M_\odot$ and yields a match $\mathcal{M} = 0.99$. The middle panel shows $M_{\rm Lz} = 10^3\,M_\odot$ with $\mathcal{M} = 0.97$, while the right panel uses $M_{\rm Lz} = 10^5\,M_\odot$ and gives $\mathcal{M} = 0.7$. These examples illustrate that lensing-induced waveform distortions can significantly enhance the signal amplitude.

stretch. The lens parameters, source distances, and the corresponding numbers of recovered injections for all sets are summarized in Table 2.1. By systematically varying lens parameters and source distances across these eleven sets, the study separates the impact of lensing induced waveform distortions from that of changing signal strength on matched-filtering detectability.

2.2.2.1 Injection campaign 1: The role of lensing

The first injection campaign focuses on how lensing amplification affects GW detectability. Four injection sets are used: one nonlensed set and three lensed sets with fixed impact parameter $y = 0.01$ and lens masses $M_{\rm Lz} = \{10, 10^3, 10^5\}\,M_\odot$. These choices span different lensing regimes, from the long wavelength limit with almost no amplification, through an intermediate regime with strong magnification but minimal beating, to a wave dominated regime with pronounced interference patterns, as depicted in Fig. 2.2.

As the lensing effect increases from one set to the next, the match between the lensed and nonlensed waveforms decreases, with typical values $\mathcal{M} = \{0.99, 0.98, 0.7\}$. In the standard optimal SNR picture, stronger lensing yields larger optimal SNRs and should therefore improve detectability. This expectation is directly assessed in the first injection campaign, with the results summarized in Fig. 2.1.

2.2.2.2 Injection campaign 2: The role of signal strength

The first injection campaign uses a fixed nonlensed SNR of about $\rho \sim 100$, representing a loud signal for O3 sensitivity and thereby minimizing the role of background noise. Although no signal of such high loudness was detected during the O3 observing run, a gravitational wave event with a comparable signal to noise ratio of approximately $\rho \sim 75$ was reported in O4 [146], highlighting the realistic possibility that similarly loud events can occur. To investigate how intrinsic signal strength alters the influence of lensing on detectability, a second injection campaign is carried out. In this campaign, attention is restricted to the lens configuration that generates the strongest waveform distortions, with $y = 0.01$ and $M_{\mathrm{Lz}} = 10^5 M_{\odot}$, which is then compared directly to the nonlensed case while the source distance is varied.

Concretely, luminosity distances $d_L = \{125, 625, 1250\}$ Mpc are considered, producing nonlensed optimal SNRs of approximately $\rho_{\mathrm{opt}} \sim \{100, 50, 10\}$ at O3 sensitivity. These setups correspond to injection set 1 and sets 4-8 in Table 2.1. This second campaign also sheds light on whether enhanced detector sensitivity, which leads to larger observed SNRs, can improve the detectability of lensed GW signals.

2.2.2.3 Injection campaign 3: detectability of lensed GWs

The first two injection campaigns separately reveal how lensing and intrinsic signal strength shape detectability, enabling more global predictions for lensed GW searches. To obtain a wider perspective on how lens parameters impact matched-filtering searches, the third campaign draws injections uniformly over $y \in [0.01, 1]$ and $M_{\mathrm{Lz}} \in [10\, M_{\odot}, 10^5\, M_{\odot}]$ for each injection set. In addition, optimal SNR based expectations are tested by directly comparing the matched-filter SNRs of lensed and nonlensed signals at luminosity distances $d_L = \{125, 625, 1250\}$ Mpc, corresponding to injection sets 1, 5, 6, and 9–11 in Table 2.1. These six sets also probe whether it may be necessary to extend existing template banks to include lensed waveforms, by examining detectability across a range of SNRs in matched-filtering searches.

In the first two campaigns (sets 1–8), the component masses, spins, lens parameters, and source distances are held fixed within each injection set. Injecting these signals at different times in the data allows a controlled comparison of the recovered chirp mass $\mathcal{M}_c$ and effective spin χ_{eff} under varying noise realizations, directly illuminating how lensing and signal strength affect the matched-filter SNR. In contrast, the third campaign samples over lens parameters, sacrificing this degree of control in favor of a more generic characterization of lensing impacts on detectability, since each injection experiences a distinct combination of noise and lens configuration. Consequently, recovered $\mathcal{M}_c$ and χ_{eff} are reported only for the first two campaigns, whereas matched-filter SNR, the signal-consistency test value ξ^2, and the false-alarm rate are presented for all three campaigns.

2.3 Results

These results show that gravitational lensing can substantially modify both the matched-filter SNR and the signal-consistency test value ξ^2. The size of the impact is controlled jointly by the level of lensing amplification and by the intrinsic strength of the signal, and therefore cannot be inferred from optimal

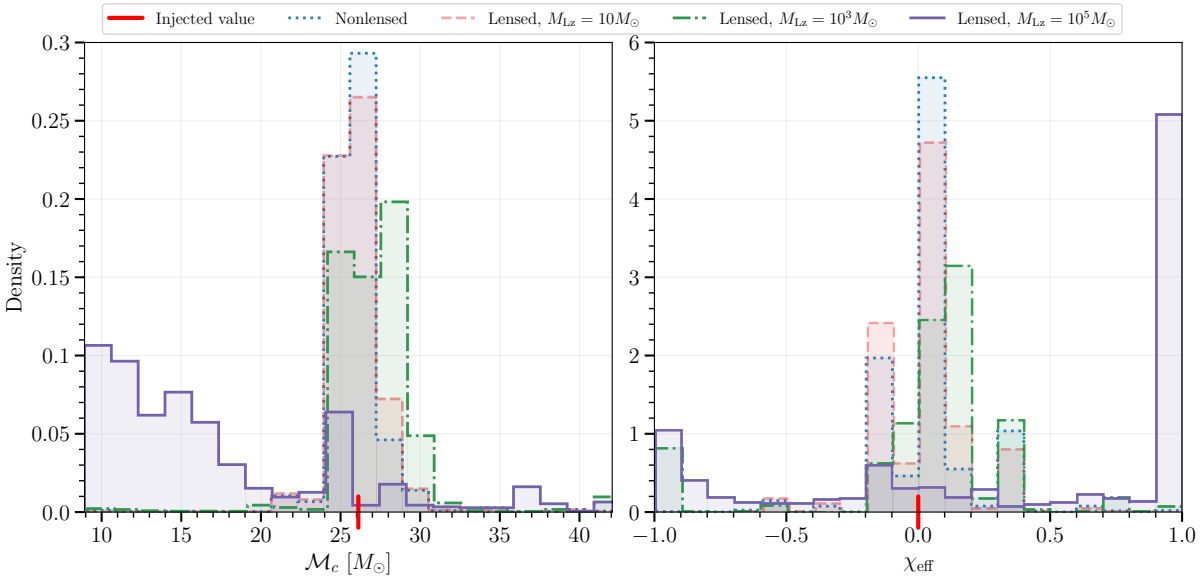

Figure 2.3: Impact of lensing on the accuracy of recovered source parameters for lensed gravitational waves. The panels show histograms of inferred source parameters for different lens configurations from the first injection campaign described in Table 2.1 and Fig. 2.2. The left panel shows the recovered chirp mass $\mathcal{M}_c$, and the right panel shows the recovered effective spin $\chi_{\rm eff}$. The red vertical line marks the injected value corresponding to a nonspinning binary black hole with component masses of 30 $M_\odot$ at a luminosity distance of 125 Mpc. Nonlensed injections are shown in blue. Lensed injections have impact parameter $y = 0.01$ and redshifted lens masses $M_{\rm Lz} = 10\,M_\odot$ with match $\mathcal{M} = 0.99$ (red), $10^3\,M_\odot$ with $\mathcal{M} = 0.97$ (green), and $10^5\,M_\odot$ with $\mathcal{M} = 0.7$ (purple). These results show that lensing-induced waveform distortions can substantially degrade the accuracy of the recovered source parameters.

SNR based arguments alone. In the following sections, these effects are analyzed in detail, leading to a comprehensive evaluation of the detectability of lensed GW signals across the explored parameter space.

2.3.1 The role of lensing

The recovered distributions of the chirp mass $\mathcal{M}_c$ and effective spin $\chi_{\rm eff}$ for sets one through four are shown in Fig. 2.3. At a luminosity distance of 125 Mpc, the nonlensed injections yield an expected SNR of order $\rho \sim 100$ at O3 sensitivity, so noise has only a modest influence on the inferred template parameters. The nonlensed set (blue) and set 2 (red) display nearly identical distributions, both peaked at $\mathcal{M}_c = 26.7\,M_\odot$ and $\chi_{\rm eff} = 0$. For the configuration with $M_{\rm Lz} = 10^3\,M_\odot$ (green), the recovered $\chi_{\rm eff}$ shifts toward values around 0.25, while the $\mathcal{M}_c$ distribution remains broadly similar to that of the nonlensed case. By contrast, set 4 (purple) experiences large lensing amplification, producing a marked reduction in the recovered $\mathcal{M}_c$ and a multimodal $\chi_{\rm eff}$ distribution with peaks near -1, 0, and $+1$. This demonstrates that even at SNR of ~ 100, significant lensing-induced waveform distortions can substantially bias the recovered source parameters and thereby lower the matched-filter SNR. At the same time, the right panel of Fig. 2.1 shows that lensing magnification tends to increase the SNR, so the resulting matched-filter SNR reflects a competition between parameter bias, which suppresses ρ, and amplification, which enhances it. This interplay is illustrated in Fig. 2.4, which plots ξ^2/ρ^2 versus ρ for sets one to four, with the false-alarm

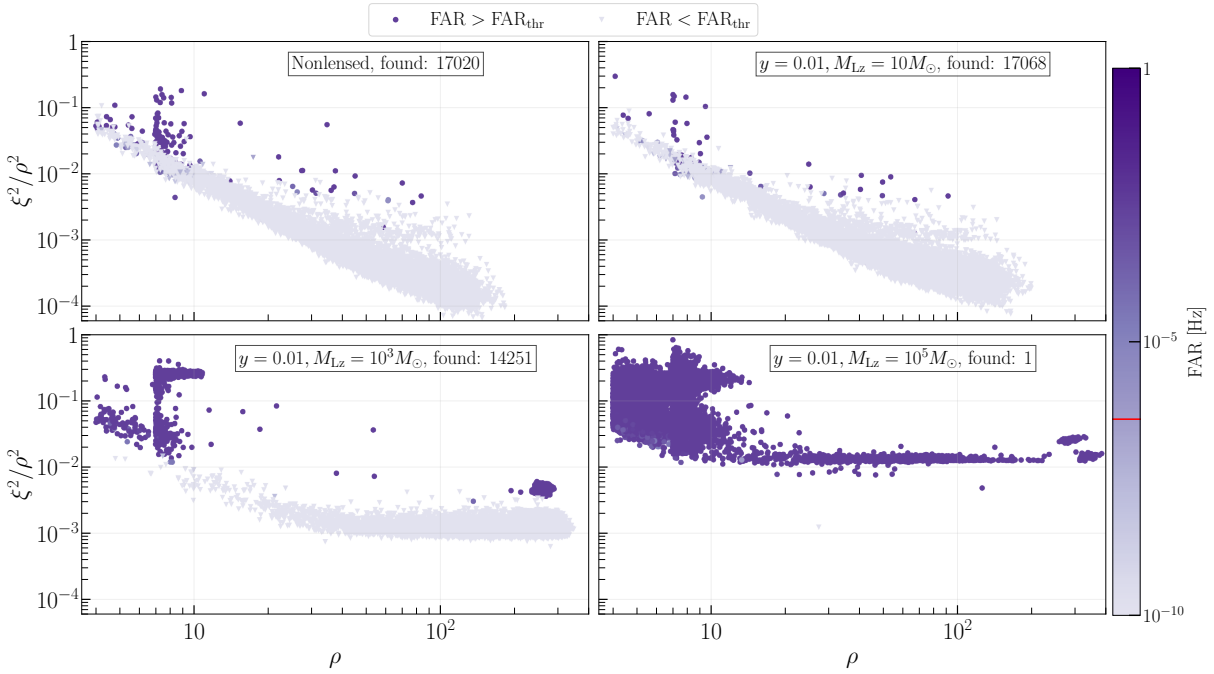

Figure 2.4: Role of lensing in the detectability of lensed GW. The panels show the distribution of simulated events from the first injection campaign described in Table 2.1 and Fig. 2.2, plotted as a function of the matched-filter SNR ρ and the signal-consistency statistic ξ^2/ρ^2 defined in Eq. 1.49. The color scale encodes the false-alarm rate, and an event is classified as a significant GW candidate when its false-alarm rate drops below 3.85×10^{-7} Hz (1 in 30 days), indicated by the red tick. Circular markers denote events above this detection threshold, while triangular markers indicate events below it. The upper-left panel shows the nonlensed injections, and the remaining panels show lensed injections with impact parameter $y = 0.01$: the upper-right panel corresponds to a redshifted lens mass $M_{\rm Lz} = 10\,M_\odot$ with match $\mathcal{M} = 0.99$, the bottom-left to $M_{\rm Lz} = 10^3\,M_\odot$ with $\mathcal{M} = 0.97$, and the bottom-right to $M_{\rm Lz} = 10^5\,M_\odot$ with $\mathcal{M} = 0.7$. Lensing-induced waveform distortions can substantially reduce ρ and increase ξ^2 in matched-filter searches, thereby decreasing the detectability of some lensed events.

rate encoded in the color scale. In the nonlensed case (upper left panel), larger SNR and smaller ξ^2/ρ^2 correlate with lower false-alarm rates, and 17,020 of 19,432 injections exceed the detection threshold of false-alarm rate $\mathrm{FAR}_{\rm thr} = 3.85 \times 10^{-7}$ Hz. Even for this simple nonlensed, nonspinning 30–30 $M_\odot$ binary at fixed distance, sampling over sky position, orientation, and noise realizations produces a wide spread of recovered SNR values, roughly from $\rho \sim 10$ up to $\rho \sim 100$, already indicating that the matched-filter SNR can differ significantly from the optimal SNR. The weakly lensed case (upper right panel) shows a very similar pattern, with a slightly higher detection count of 17,068, reflecting mild amplification and negligible waveform distortion.

The intermediate lensing configuration (lower left panel) highlights the competing roles of imperfect parameter recovery and magnification. Overall, the SNR distribution shifts to higher values due to lensing amplification, but there is also a noticeable rise in the number of injections recovered with $\rho < 10$. Consequently, the achieved SNR becomes highly sensitive to both the specific noise realization and the sampled source parameters. After computing ρ, the nonlensed template used to recover the lensed injection defines an expected SNR time series, while the actual SNR time series incorporates lensing

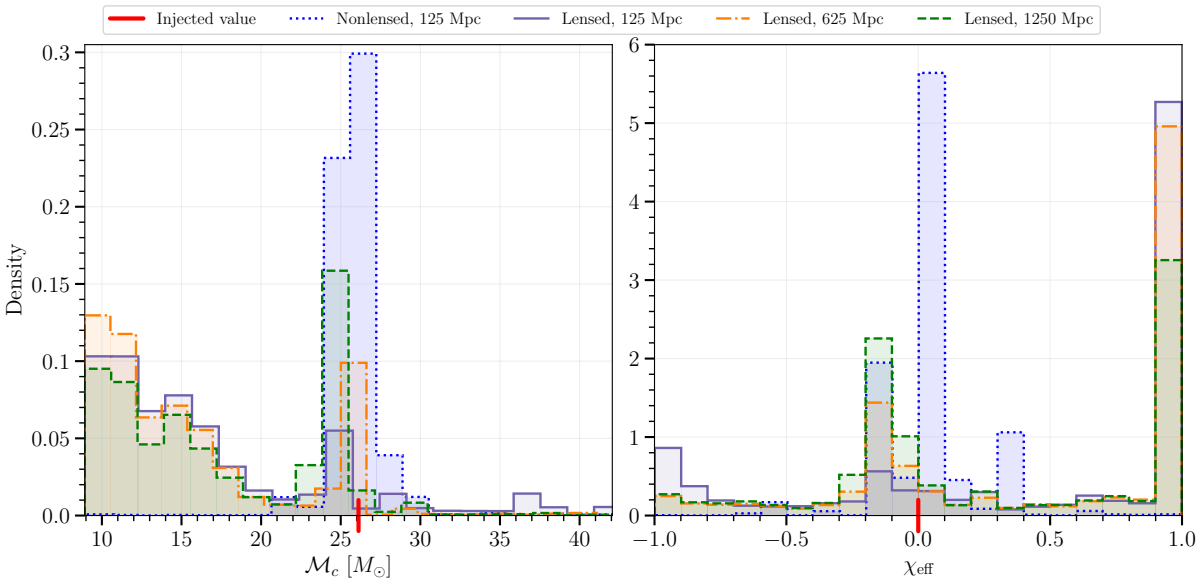

Figure 2.5: Role of signal strength in the accuracy of recovered source parameters for lensed GW. The panels show histograms of recovered source parameters at different source distances from the second injection campaign described in Table 2.1. The left panel displays the chirp mass $\mathcal{M}_c$, and the right panel shows the effective spin $\chi_{\rm eff}$. The red vertical line marks the injected value corresponding to a spinless binary black hole with component masses of $30\,M_\odot$. Nonlensed injections at a distance of 125 Mpc are shown in blue. All lensed injections use an impact parameter $y = 0.01$ and a redshifted lens mass $M_{\rm Lz} = 10^5\,M_\odot$, with source distances of 125 Mpc (purple), 625 Mpc (orange), and 1250 Mpc (green). When lensing-induced distortion is significant, the recovered source parameters become more accurate at larger source distances, corresponding to lower SNR, because the down-weighting of highly distorted high-SNR events reduces systematic biases in the parameter estimation.

induced distortions. When ξ^2 is evaluated using Eq. 1.49, these distortions appear as elevated ξ^2 values. Although ξ^2 is designed to downweight loud noise artifacts, it does not distinguish between deviations caused by noise and those arising from missing physical effects in the waveform model. As a result, ξ^2/ρ^2 increases overall, with the largest enhancement at high SNR, rising from typical values $\sim 10^{-4}$ to $\sim 10^{-3}$. Because both ρ and ξ^2 enter the false-alarm rate, this shift drives the false-alarm rate upward and reduces the number of detections to 14, 251.

The most strongly-lensed set represents an extreme scenario in which only 1 of 19, 421 injections is successfully detected. A large fraction of these lensed signals are recovered with optimal SNR below $\rho \sim 10$ due to severe mismatches between the true waveform and the best fitting template. The poorly matched template then generates an SNR time series that, when used in the computation of ξ^2, produces false-alarm rate outside the range for which the statistic was designed. This pushes the false-alarm rate from below 10^{-10} to above 1 Hz, effectively eliminating all but one injection.

Taken together, these findings show that strong waveform distortions can dramatically alter the behavior of the matched-filter SNR, implying that the optimal SNR is a poor surrogate for detectability even when $\rho_{\rm opt} \sim 100$. Furthermore, employing a very large template bank with millions of templates in a realistic search increases the coverage of parameter space but also broadens the distribution of recovered source parameters. This wider recovered space tends to yield less accurate SNR time series for use in ξ^2,

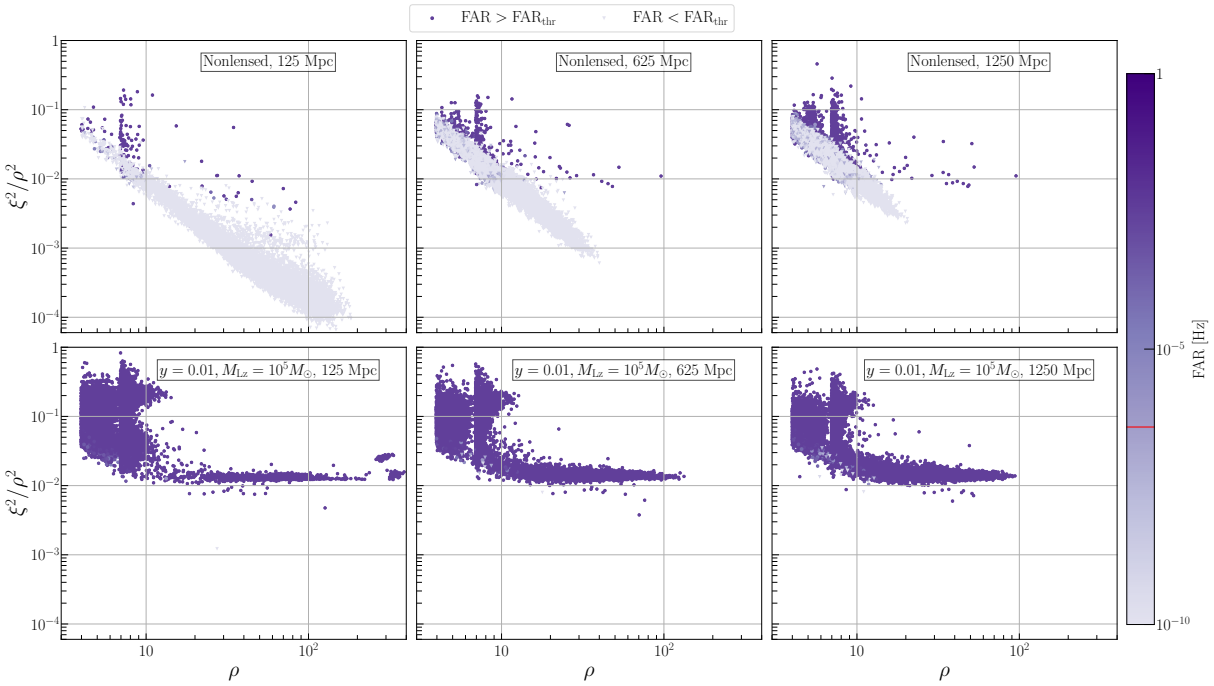

Figure 2.6: Role of signal strength in the detectability of lensed GW. The panels show the distribution of simulated events from the second injection campaign described in Table 2.1, plotted as a function of the matched-filter SNR ρ and the signal-consistency statistic ξ^2/ρ^2 defined in Eq. 1.49. The color scale indicates the false-alarm rate. An event is classified as a significant GW candidate when its false-alarm rate lies below 3.85×10^{-7} Hz (1 in 30 days), marked by the red tick. Circular markers denote events above this detection threshold, while triangular markers denote events below it. The upper panels show nonlensed injections. The lower panels show lensed injections with impact parameter $y = 0.01$ and redshifted lens mass $M_{\rm Lz} = 10^5\, M_\odot$: the left panel corresponds to a source distance of 125 Mpc, the middle to 625 Mpc, and the right to 1250 Mpc. These results demonstrate that GWs with strong wave-optics effects are less detectable at smaller source distances, where the SNR is higher, because the distortion drives up ξ^2 and causes such events to be down-ranked in matched-filtering searches.

degrading the reliability of the statistic and further complicating detectability assessments. The outcomes of this first injection campaign therefore also support the use of a smaller, carefully chosen template bank in detectability studies.

A fully comprehensive explanation of the biases in the recovered parameters for lensed injections would require a detailed, template by template analysis of the response, including inspection of the SNR and ξ^2 time series for individual events. Such an investigation would be computationally intensive and is left to future work.

2.3.2 The role of the signal strength

Figure 2.5 contrasts the recovered chirp mass $\mathcal{M}_c$ and effective spin $\chi_{\rm eff}$ for nonlensed injections at 125 Mpc (blue) with strongly-lensed injections at 125 Mpc (purple), 625 Mpc (orange), and 1250 Mpc (green). At every distance, the lensed sets show clear departures from the nonlensed distributions, demonstrating that strong wave-optics effects significantly reduce the accuracy of recovered source parameters once

detector noise is included. As the luminosity distance increases, the discrepancy between lensed and nonlensed distributions shrinks, indicating that in noisier data matched-filtering searches becomes less sensitive to large waveform mismatches and therefore suffers a smaller loss in matched-filter SNR than in the first campaign with very loud signals.

Figure 2.6 compares detectability for nonlensed injections (upper panels) and for lensed injections with $y = 0.01$ and $M_{\mathrm{Lz}} = 10^5 M_\odot$ (lower panels) at 125 Mpc (left), 625 Mpc (middle), and 1250 Mpc (right). For nonlensed injections, increasing distance systematically lowers the SNR and raises ξ^2, with 17,020, 12,878, and 6,217 injections recovered at 125, 625, and 1250 Mpc, respectively. In contrast, the lensed injections exhibit the opposite trend: as the distance grows and the intrinsic SNR decreases, a larger fraction of lensed signals is recovered with higher matched-filter SNR than in the corresponding nonlensed case, showing that the SNR loss caused by lensing distortions is partially mitigated in noisier backgrounds.

For the lensed sets at all distances, detectability is controlled primarily by the behavior of the signal consistency statistic rather than by SNR itself. Elevated false-alarm rates are driven mainly by large ξ^2 values produced by lensing induced waveform distortions, with only a weak dependence on the actual SNR. As a result, only 1, 8, and 13 lensed injections are recovered at 125, 625, and 1250 Mpc, respectively, despite magnification. This campaign therefore shows that matched-filtering searches tend to reject signals that deviate strongly from the template family more aggressively at higher SNR, so improved detector sensitivity does not automatically translate into better detectability for strongly distorted lensed signals.

2.3.3 General picture of detectability of lensed GWs in matched-filtering searches

These results summarize how lensing across a broad parameter range alters matched-filter SNR, ξ^2, and false-alarm rate, and why optimal SNR alone mischaracterizes detectability. The discussion explains where lensing reduces SNR, how ξ^2 responds, and which lens parameters most strongly degrade detection performance.

Figure 2.7 displays the ratio of matched-filter SNR for lensed versus nonlensed injections together with the corresponding ξ^2 values and false-alarm rate across the (y, M_{Lz}) plane. In contrast with naive optimal SNR expectations, which would suggest $\rho_{\mathrm{L}}/\rho_{\mathrm{NL}} > 1$ everywhere, the upper left panel reveals a purple region where $\rho_{\mathrm{L}}/\rho_{\mathrm{NL}} < 1$. This arises from the competition between lensing induced waveform distortion, which biases parameter recovery and suppresses the SNR, and lensing amplification, which boosts the amplitude and hence the SNR. The purple band therefore marks the part of parameter space in which parameter recovery errors dominate over magnification, and it coincides with the region where the match satisfies $\mathcal{M} \lesssim 0.75$ in Fig. 2.1, hinting at a connection between match and recovery bias.

Outside this region, the SNR ratio agrees well with optimal SNR predictions, corresponding to the domain in Fig. 2.1 where amplification dominates over parameter recovery errors. Comparing the upper left and upper right panels of Fig. 2.7 shows that the area with $\rho_{\mathrm{L}}/\rho_{\mathrm{NL}} < 1$ shrinks as the source distance increases. This trend matches the second injection campaign: at very high SNR around $\rho \sim 100$, lensing distortion strongly degrades recovery, whereas at lower intrinsic SNR around $\rho \sim 10$, background noise partially masks the distortion, leading to a smaller effective SNR loss.

The middle left panel of Fig. 2.7 shows how ξ^2 varies with (y, M_{Lz}) at 125 Mpc. Regions with strong

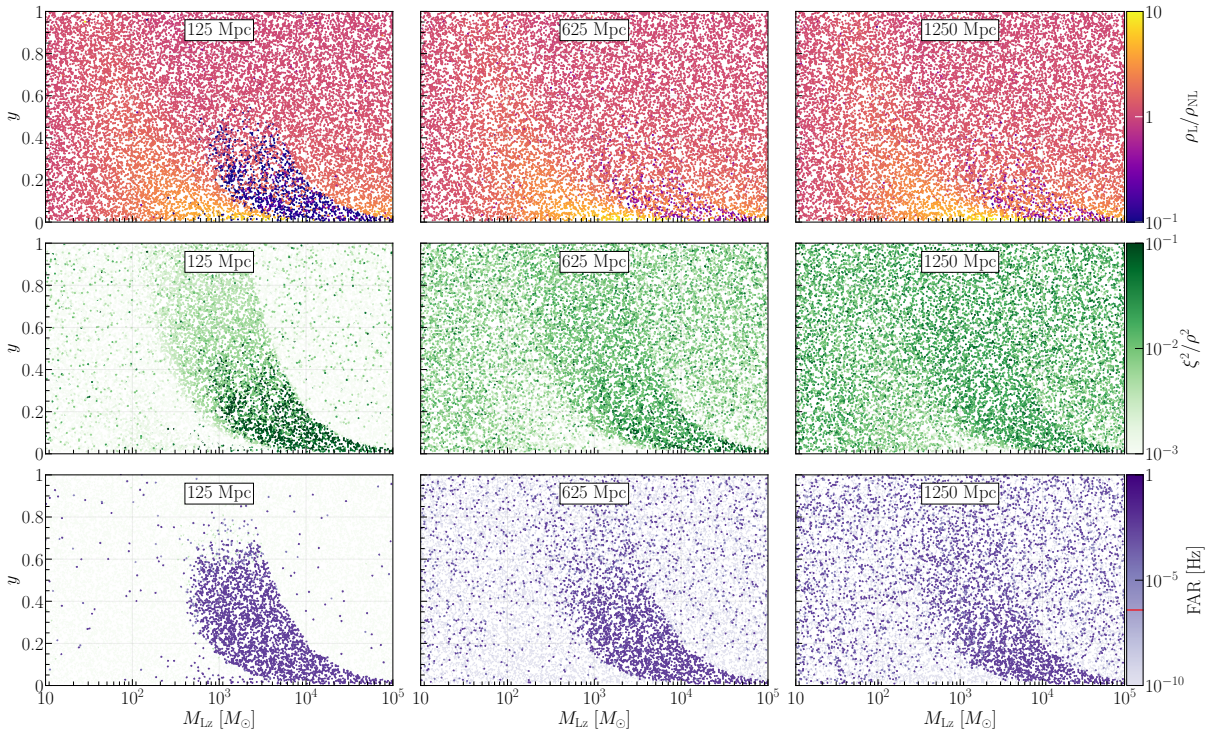

Figure 2.7: Detectability of lensed GWs. The panels show the distribution of simulated events from the third injection campaign described in Table 2.1, plotted in the plane of impact parameter y and redshifted lens mass M_{Lz}. From top to bottom, the color scale represents the ratio of the lensed matched-filter SNR ρ_L to the nonlensed SNR ρ_{NL}, the signal-consistency statistic ξ^2/ρ^2 defined in Eq. 1.49, and the false-alarm rate. An event is classified as a significant GW candidate when its false-alarm rate is below 3.85×10^{-7} Hz (1 in 30 days), indicated by the red tick. The source distances shown from left to right are 125, 625, and 1250 Mpc. These results show that a substantial region of the lens-parameter space fails to satisfy the false-alarm-rate threshold, even where the SNR is high, because strong wave-optics effects increase the signal-consistency statistic and cause events to be down-ranked in matched-filtering searches.

wave-optics effect exhibit large ξ^2 values, and these regions overlap closely with those where $\rho_L/\rho_{NL} < 1$ in the upper panel. This overlap indicates that the same parts of parameter space are responsible both for SNR suppression and for inflated ξ^2, and hence for higher false-alarm rate. As the distance increases, ξ^2 generally grows because of stronger background noise, illustrating how noise can partially hide waveform distortions while still worsening the overall ranking statistics.

The bottom left panel of Fig. 2.7 confirms that regions with large false-alarm rate coincide with areas of reduced SNR and elevated ξ^2 seen in the upper and middle panels. This correlation demonstrates that SNR loss and increased ξ^2 due to strong wave-optics effect act together to render lensed signals difficult to detect. Moving from the bottom left to the bottom right panels, the fraction of lensed injections with false-alarm rate above threshold increases overall, as weaker signals in noisier data produce less significant ranking statistics once ρ, ξ^2, and the other terms in Eq. 1.51 are combined. At the same time, the area with extremely large false-alarm rate shrinks with distance, in line with the SNR and ξ^2 trends.

The third injection campaign leads to several key conclusions. First, optimal SNR based analyses miss the substantial SNR loss generated by lensing distortions combined with signal consistency tests, especially

in wave-optics regimes. As a consequence, lensing studies and rate estimates that rely only on optimal SNR tend to overestimate detectability for strongly-lensed signals. Second, despite an overall drop in detectability, signals that deviate most strongly from nonlensed templates are, somewhat counterintuitively, more easily recovered at lower intrinsic SNR because background noise partially mitigates the impact of waveform mismatches on the ranking statistic. This implies that improved detector sensitivity does not automatically translate into better recovery of strongly-lensed events. Finally, Fig. 2.7 underscores the need for dedicated lensed waveform template banks and points to the specific regions in (y, M_{Lz}) space that are most important for future template construction, particularly in the point mass lens scenario.

2.4 Conclusions

This work shows that the detectability of new classes of GW signals in matched-filtering searches depends sensitively on both template completeness and the detector noise background. When key physical effects are absent from the template bank, the resulting waveform mismatches can severely suppress detection efficiency. A prime example is gravitational lensing in the wave-optics regime, where interference induced beating features distort the waveform morphology in ways not captured by nonlensed templates [26]. Optimal SNR is often used as a proxy for detectability in this context, but it ignores the possibility that detector noise can mimic or deform real signals. For this reason, modern search pipelines incorporate additional statistics, such as signal consistency tests, to reject noise artifacts, making an accurate characterization of the noise background essential for assessing the astrophysical nature of candidates.

In this study, the detectability of lensed GW signals is examined systematically for the first time using a state of the art matched-filtering search pipeline that includes realistic ranking statistics and real detector noise. Simulated lensed signals are injected into data containing actual noise and are recovered with the time domain matched-filtering search pipeline `GstLAL` [23, 41, 147, 148]. Three injection campaigns are carried out to quantify detectability. The first two isolate, respectively, the roles of lensing and intrinsic signal strength, while the third maps detectability across the lens parameter space. In the first two campaigns, nonlensed template banks can recover lensed, magnified signals with higher matched filter SNR for a fixed impact parameter, across a range of lens masses and source distances. At the same time, a non negligible fraction of lensed signals suffers significant SNR loss because the search pipeline recovers biased source parameters, and this fraction grows with both the strength of lensing amplification and the intrinsic SNR. A striking case arises for strongly magnified, distorted signals with $y = 0.01$ and $M_{\mathrm{Lz}} = 10^5 M_\odot$ at 125 Mpc: although the nonlensed SNR is of order 100, many events are recovered with matched filter SNR below 10.

The autocorrelation-based signal-consistency statistic ξ^2 plays an equally crucial role in governing detectability. When lensed signals are searched using an nonlensed template bank, strong waveform distortions drive ξ^2 to large values, causing it to dominate the ranking statistic. The combined effect of SNR loss and elevated ξ^2 can boost the false-alarm probability by many orders of magnitude. For example, the detection efficiency at 125 Mpc drops from about 90% in the nonlensed case to below 1% for $y = 0.01$ and $M_{\mathrm{Lz}} = 10^5 M_\odot$. These effects lie entirely outside the reach of optimal SNR based analyses

and demonstrate that lensed GWs are generally less detectable in matched-filtering searches than optimal SNR thresholds would suggest. Despite the drop in detection efficiency of lensed GWs in a time-domain matched-filtering search pipeline `GstLAL`, other frequency domain matched-filtering pipelines, such as `PyCBC` [149], MBTA [150, 151] and `IAS-HM` [152]

The third injection campaign shows that lensed signals in the wave optics regime, which produce the largest waveform distortions, deviate most strongly from optimal SNR predictions. This identifies the wave optics regime as the parameter space region where lensing has the most severe impact on detectability and thus where dedicated lensed template banks are most urgently needed. Motivated by these findings, several strategies for detecting lensed GWs are proposed. First, matched-filtering searches that employ lensed template banks are required to robustly capture lensing signatures, especially in regions of strong morphological distortion. Building such banks will demand new injection campaigns that sample the full joint source–lens parameter space, as well as careful evaluation of the computational costs associated with increasing the dimensionality of the bank. Second, coherent wave burst pipelines, which are largely model independent, offer a complementary avenue [153]. Unlike template based matched-filtering search, coherent burst methods can identify loud signals without assuming a specific waveform model, making them potentially more sensitive to lensing induced modifications; targeted lensing analyses of candidates found only by such searches provide an alternative route to discovering lensing signatures missed by standard matched-filtering search pipelines.

The gap between detectability in full matched-filtering searches and predictions based solely on optimal SNR has important implications for broader lensing studies. In particular, the detectability of lensed signals in the wave dominated regime is substantially overestimated when nonlensed searches are approximated using optimal SNR thresholds. Follow up parameter estimation efforts that seek wave optics signatures necessarily draw from existing catalogs, which are unlikely to include strongly distorted lensed events, leading to selection biases. Correcting these biases requires large scale injection campaigns that vary both source and lens parameters to determine detection probabilities at the parameter estimation stage. Likewise, lensing forecasts and constraints that rely on optimal SNR must be revisited. Current rate estimates for lensed events and constraints on lens populations, especially compact dark matter lenses in the wave dominated regime, typically use optimal SNR criteria [154, 76, 154, 47, 69, 74, 75, 155, 156, 80, 157]. Because they neglect the reduced detectability demonstrated here, these analyses tend to be overly optimistic in parts of parameter space, and a comprehensive reanalysis based on full injection campaigns is needed.

More generally, searches for new physics in GW data may require dedicated detection pipelines rather than simple optimal SNR extrapolations. This study provides a concrete example in which optimal SNR arguments fail once new physical effects. Here, gravitational lensing in the wave optics regime are absent from the template bank. Missing physics alters both recovered source parameters and signal consistency statistics, thereby degrading detectability even for intrinsically loud signals. The methodology developed here is broadly applicable and can be extended to other forms of beyond standard waveform physics. Future work will include additional injection campaigns targeting, for example, deviations from general relativity, to explore how such effects modify detectability in the presence of realistic noise. These studies will help clarify the interplay between detector noise and unmodeled physics in matched-filtering searches

and may guide the design of more robust detection strategies for a wide range of novel GW signals.

Chapter 3

Discovering distorted gravitational waves by lensing using deep learning

In the previous chapter, we demonstrated the difficulty of detecting lensed gravitational waves in search pipelines. This highlights the need for lensing dedicated matched filtering searches and for placing greater emphasis on results from coherent-wave burst searches. However, this challenge is not the final obstacle to the discovery of lensed GWs in wave-optics regime. After a lensed GW candidate is identified in a search, one must perform inference under both lensed and unlensed hypotheses to assess the statistical evidence for lensing. Establishing robust significance typically requires thousands to millions of inference evaluations in order to disentangle waveform systematics from noise fluctuation and to reach three sigma or five sigma confidence. Using traditional sampling algorithm, which can require on $\sim$ 50 CPU days, this task becomes computationally prohibitive. To overcome this bottleneck and enable the discovery of lensed GWs, this chapter introduces a simulation-based inference framework, `DINGO-lensing`, that dramatically reduces the computational cost of parameter inference and model comparison, while maintaining similar accuracy with traditional sampling methods. As an application, we also utilize simulation-based inference to throughoutly analyze GW231123 and assess the statistical significance.

This chapter is structured as follows: Setction 3.1 summarizes the current efforts of lensed GW inferernce. Section 3.2 describes the lensed waveform model, network architecture, and the training setup with O5 sensitivity and for GW231123 analysis. Section 3.3 presents the verification of `DINGO-lensing` and details the GW231123 analysis. Section 3.4 summarizes our findings and discusses the implications for future discovery of simulation-based inference in lensing analyses.

3.1 Current efforts of gravitational lensing inference

Confirming that a candidate event is lensed requires detailed scrutiny and sensitivity to subtle effects, including noise fluctuations and waveform systematics [138]. Achieving a robust detection claim typically involves simulating and analyzing very large GW event samples, often reaching into the millions. By contrast, the fourth observing run produced a lensed candidate, GW231123 [131], which is investigated in detail in a companion analysis [3].

A central challenge for GW lensing studies is the substantial computational cost of performing Bayesian inference on distorted signals using standard sampling schemes such as MCMC and nested sampling [24, 158]. This expense both constrains the ability to keep pace with the rapidly growing detection rate expected from improved detector reach and makes comprehensive background estimation for lensing searches extremely difficult. In particular, determining a lensing false-alarm rate would require millions of nonlensed parameter estimation runs. To alleviate these issues, several alternative strategies have been proposed that avoid conventional Bayesian inference [159, 160]. While these approaches can deliver significant computational speedups, they generally do so at the cost of the full statistical interpretability and robustness provided by traditional Bayesian methods.

Deep learning techniques have emerged as a powerful tool for rapid and accurate parameter estimation of GW signals [158]. Unlike traditional Bayesian samplers, which require repeated waveform generation and are therefore computationally expensive, neural networks are trained to learn the highly nonlinear mapping from GW strain data to source parameters directly from large sets of simulated examples. Once trained, these models can return full posterior distributions within seconds, achieving speedups of several orders of magnitude while maintaining accuracy comparable to standard parameter estimation pipelines [161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173].

Recent work has begun to adapt deep learning methods for the rapid identification of strongly lensed GW signals [174, 175]. In the wave optics regime, a range of architectures has been investigated, including visual geometry group style convolutional networks [176], conditional variational autoencoders [177], and neural spline flow based density estimators [160]. Reference [176] offers a proof of principle that deep models can recognize lensing induced waveform distortions, while Reference [177] demonstrates a substantial speedup in lens parameter inference relative to traditional Bayesian samplers. Building on this, Reference [160] pushes the inference problem to an eleven dimensional parameter space. Nevertheless, none of these approaches yet delivers rapid inference of the complete set of source parameters with accuracy fully comparable to that of standard Bayesian sampling techniques.

The `DINGO` framework [56, 57, 58, 59, 60, 61, 62] is a state-of-the-art implementation of simulation based inference for GW data using neural posterior estimation. By combining an embedding network with normalizing flows, `DINGO` directly models full Bayesian posterior distributions conditioned on GW observations, enabling rapid, high fidelity inference in high dimensional parameter spaces [57, 58]. Simulation-based inference alleviates the tension between computational speed and statistical rigor by avoiding explicit likelihood evaluations; instead, neural density estimators are trained on large sets of simulated data to learn the nonlinear mapping from observed strain to source parameters [178, 158]. Within this framework, `DINGO` offers a promising deep learning pathway toward fast and accurate inference for lensed GW signals, retaining the key advantages of Bayesian posteriors while dramatically reducing computational cost [57, 58].

Building on the simulation based inference framework `DINGO`, a new lensing inference pipeline, `DINGO-lensing`[1] has been developed to target lensed GW signals. This pipeline employs a flexible two image waveform model for lensed signal generation that captures the most commonly studied wave optics lensing configu-

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

rations [64]. By learning posterior distributions directly from simulated data, `DINGO-lensing` sidesteps the computational bottlenecks of traditional sampling based methods while retaining the statistical rigor of standard Bayesian inference. This strategy offers a scalable and robust avenue for identifying and characterizing lensed GW events in the face of rapidly growing detection rates. The verfication of `DINGO-lensing` is presented in Section 3.3.1

GWTC-4.0 reports GW231123 as the most massive and highly spinning binary black hole detected so far [55, 131], and it is also identified as the highest ranked single event lensing candidate in current LVK searches [5]. The LVK's detailed follow up finds GW231123 to be an outlier whose interpretation is complicated by waveform systematics, potential noise artifacts, and low prior odds for strong lensing, preventing a definitive conclusion. Independent analyses have reported strong support for lensing [132, 179], as well as alternative scenarios such as overlapping signals [179], but a fully robust, quantitative significance for the lensing hypothesis has not yet been established in the literature.

The small number of observable cycles in very massive mergers like GW231123 makes multiple competing explanations difficult to distinguish, as already seen for GW190521 [180, 181], which has been interpreted in terms of eccentric binaries [182], dynamical captures [183], head-on Proca star collisions [184], and primordial black hole scenarios [185]. Given inferred merger rate densities of order 0.1 $\mathrm{Gpc}^{-3}, \mathrm{yr}^{-1}$ for GW190521/GW231123 like events [131, 181], additional detections of this population are expected in future observing runs, underscoring the need for robust inference frameworks that can both recover source parameters reliably and quantify the statistical significance of competing physical interpretations, including lensing.

In Section 3.3.2, `DINGO-lensing` is used to carry out a comprehensive reanalysis of GW231123, quantifying the statistical significance of the lensing hypothesis by training neural posterior estimators on large ensembles of lensed and nonlensed simulations tailored to GW231123-like sources. The resulting methodology and detection statistics provide scalable strategies for identifying and validating lensed GW events in upcoming observing runs, where traditional samplers would be computationally prohibitive.

3.2 Methods

In this section, we detailed our waveform generation, lensed waveform model, training and how we obtain the lensing Bayes factor.

3.2.1 `DINGO-lensing`

Neural posterior estimation is carried out using the simulation based inference framework `DINGO`, trained on GW signals simulated with the detector sensitivity. The training assumes Gaussian, stationary noise characterized by the fifth observing run amplitude spectral density for the LVK network [186, 187]. Glitches can violate these assumptions by introducing non Gaussian, non stationary features, and their impact on the method's performance will be explored in future work.

For waveform modeling, the analysis employs the inspiral–merger–ringdown phenomenological model with higher modes and precession, `IMRPhenomXPHM` [188], as implemented in `LALSimulation` [145,

189]. Signals are generated in the frequency band [20, 1024] Hz with a duration of 8 s. Our framework is further extended to support numerical relativity surrogate waveforms, such as `NRSur7dq4`, which covers [20, 512] Hz and is used in the companion analysis to GW231123 (See Section 3.3.2 and Reference [3]).

The intrinsic parameter set includes the detector frame component masses m_i for $i = 1, 2$, spin magnitudes a_i, and spin orientation angles $(\phi_i, \phi_{12}, \phi_{JL})$ [188]. Additional extrinsic parameters are the inclination θ_{JN}, sky location (right ascension and declination), polarization angle ψ, luminosity distance $d_{\rm L}$, geocenter time of peak amplitude $t_{\rm c}$, and the phase at the reference frequency, $\phi_{\rm ref}$ evaluated at 20 Hz.

3.2.2 Lensed Waveform Model

In this work, we model GW lensing in the frequency domain using the weak gravity and thin lens approximations, where the lensed signal is described by a frequency-dependent amplification factor applied to the unlensed waveform. We focus on a two image interference model characterized by a relative magnification and a time delay with a fixed phase difference, which captures the dominant wave-optics effects relevant for point, fold, and many cusp lensing configuration. The detailed physical motivation and the assumptions of this lensing model are discussed in Section 1.4.1.6. The lens model is implemented in `modwaveforms`.[2].

3.2.3 Training setup of O5 network

GW parameter estimation is performed with `DINGO-lensing`, which builds on the simulation based inference pipeline `DINGO` [56, 57, 58, 59, 60, 61]. In this framework, neural posterior estimation is used to learn a conditional density $q_\phi(\boldsymbol{\theta}|d)$ that approximates the exact Bayesian posterior $p(\boldsymbol{\theta}|d)$, where ϕ denotes the trainable network parameters, $\mathcal{L}(d|\boldsymbol{\theta})$ is the likelihood, $\pi(\boldsymbol{\theta})$ is the prior, and $\mathcal{Z}(d)$ is the evidence. Training is carried out on simulated, labeled pairs $(\boldsymbol{\theta}, d)$: source parameters $\boldsymbol{\theta}$ are drawn from $\pi(\boldsymbol{\theta})$, and the corresponding strain data d are generated from $\mathcal{L}(d|\boldsymbol{\theta})$. In practice, each training example is obtained by injecting a waveform $h(\boldsymbol{\theta}_i)$ into stationary Gaussian noise n_i, so that $d_i = h(\boldsymbol{\theta}_i) + n_i$.

The network architecture follows the design of Reference [59], with all hyperparameters kept fixed. In brief, `DINGO-lensing` employs a two-stage setup consisting of an embedding network and a density estimator. The embedding network first compresses the raw strain time series into a 128 dimensional representation using singular value decomposition, and this low dimensional embedding is then processed by a fully connected network that parameterizes the approximate posterior density. The full model contains on the order of 10^8 trainable parameters and about 10^7 fixed parameters, providing sufficient expressivity to capture the complex, high dimensional GW posterior structure while allowing rapid inference once training is complete [58, 57].

Training datasets are built from 5×10^6 simulated waveforms, with source parameters drawn from the prior distributions listed in Table 3.1. The waveforms are stored in a singular value decomposition representation, and 2% of the samples are reserved as an independent test set to evaluate performance.

A two stage training schedule is used, with hyperparameters listed in Table 3.2. In the first phase (Stage 0),

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

Parameter	Unit	Prior for O5 network	Prior for GW231123 network	Injection
$\mathcal{M}_c$	$M_\odot$	$\mathcal{U}(10, 100)$	$\mathcal{U}(30, 180)$	49.2
q	$M_\odot$	$\mathcal{U}(1/8, 1.0)$	$\mathcal{U}(1/6, 1.0)$	0.79
m_1	$M_\odot$	$C(5, 150)$	$C(1, 1001)$	63.6
m_2	$M_\odot$	$C(5, 150)$	$C(1, 1001)$	50.4
a_1		$\mathcal{U}(0, 0.99)$	$\mathcal{U}(0, 0.99)$	0.28
a_2		$\mathcal{U}(0, 0.99)$	$\mathcal{U}(0, 0.99)$	0.45
ϕ_{12}		$\mathcal{U}(0, 2\pi)$	$\mathcal{U}(0, 2\pi)$	3.59
ϕ_1		$\sin(0, \pi)$	$\sin(0, \pi)$	1.75
ϕ_2		$\sin(0, \pi)$	$\sin(0, \pi)$	0.49
ϕ_{JL}		$\mathcal{U}(0, 2\pi)$	$\mathcal{U}(0, 2\pi)$	5.38
θ_{JN}		$\sin(0, \pi)$	$\sin(0, \pi)$	2.83
ϕ_{ref}		$\mathcal{U}(0, 2\pi)$	$\mathcal{U}(0, 2\pi)$	1.73
RA		$\mathcal{U}(0, 2\pi)$	$\mathcal{U}(0, 2\pi)$	3.01
DEC		$\cos(-\pi/2, \pi/2)$	$\cos(-\pi/2, \pi/2)$	-0.13
ψ		$\mathcal{U}(0, \pi)$	$\mathcal{U}(0, \pi)$	2.63
t_c	s	$\mathcal{U}(-0.1, 0.1)$	$\mathcal{U}(-0.1, 0.1)$	-0.0565
d_L	Gpc	$\mathcal{U}(1, 15)$	$\mathcal{U}(0.5, 10)$	9.84
Δt	s	$\mathcal{U}(0, 0.1)$	$\mathcal{U}(0, 0.1)$	0.031
μ_{rel}		$\mathcal{U}(0, 1)$	$\mathcal{U}(0, 1)$	0.79

Table 3.1: The first four columns specify the prior distributions used for each parameter in the neural posterior estimation training set, where $\mathcal{U}$ denotes a uniform prior and C indicates that the parameter is restricted to the quoted interval. The final column lists the injected parameter values employed to evaluate the accuracy of `DINGO-lensing` and `bilby` in Figs. 1.5 and 3.4.

Training setting	Stage 0	Stage 1
Layers	SVD layer frozen	All layers unfrozen
Epochs	300	150
Optimizer	Adam [190]	Adam
Learning rate	7×10^{-5}	1×10^{-5}
Scheduler	Cosine scheduler [191]	Cosine scheduler
Batch size	4096	4096
Temperature	300	150
Purpose	Stabilize early training	Fine-tuning for accuracy

Table 3.2: Summary of the two-stage neureal posterior estimation training scheme.

the singular value decomposition layer is frozen, and it is only unfrozen during the subsequent fine tuning phase (Stage 1). Both the lensed and nonlensed neural posterior estimation networks are optimized under this two-stage scheme. In Stage 0, the model is trained for 300 epochs with the Adam optimizer [190], using a learning rate of 7×10^{-5} together with a cosine learning rate schedule [191]. During Stage 1, all layers are made trainable and the model is further fine tuned for 150 epochs with a reduced learning rate of 1×10^{-5}. In both stages, a batch size of 4096 is used, with temperatures of 300 and 150, respectively. This staged procedure enhances optimization stability in the early phase and enables the network to refine the learned embedding, which in turn yields more accurate posterior estimates.

Model performance during training is monitored via a loss function that quantifies the mismatch between the predicted and target probability distributions. Concretely, the loss is given by the negative log likelihood associated with the forward Kullback–Leibler divergence between the exact posterior $p(\boldsymbol{\theta}|d)$ and the neural approximation $q_\phi(\boldsymbol{\theta}|d)$[57]. Minimizing this objective encourages the model to reproduce posteriors that are consistent with exact Bayesian inference. Optimization proceeds epoch by epoch, updating the parameters ϕ with stochastic gradient descent–based methods until convergence, as depicted in Fig. 3.1.

Additional training setups with modified batch sizes, temperatures, and learning rates were also investigated. Although the validation results in Section 3.3.1.2 focus on the configuration given in Table 3.2, changes in batch size and temperature were found to have only minor effects on performance. In contrast, using a smaller learning rate improves training stability and leads to better performance when the luminosity distance range is extended.

3.2.4 Training setup of GW231123 network

The network architeture for GW231123 network follow Section 3.2.3 and only the difference is detailed in this section.

Neural networks are trained using the numerical relativity surrogate waveform model `NRSur7dq4` [192], over the frequency band $[20, 512]$ Hz and an analysis duration of 8 s. Among available waveform models, `NRSur7dq4` yields the lowest mismatch against numerical relativity simulations tailored to GW231123 like systems, and the starting frequency is set to 0 Hz so that the model generates the signal as early before merger as allowed by the waveform length.

The networks are trained over chirp masses in $[30, 180]M_{\odot}$, mass ratios in $[1/6, 1]$, and luminosity distances in $[0.5, 10]$ Gpc, with the remaining source parameters drawn from the priors in Section 3.2 (See the fourth column in Table 3.1), but with extended bounds on the component masses to encompass GW231123. An average LIGO noise power spectral density around the epoch of GW231123 is used to generate Gaussian noise realizations for training [193].

3.2.5 Importance Sampling

To recover accurate posterior estimates without having to retrain the neural network for every change in the likelihood configuration, `DINGO` uses an importance sampling correction step applied after the neutral posterior estimation stage [61]. Given a trained model that provides an approximation $q_\phi(\boldsymbol{\theta}|d)$ to the posterior over source and lensing parameters $\boldsymbol{\theta}$ conditioned on the data d, samples drawn from this approximate distribution are subsequently reweighted to reconstruct the desired posterior associated with a specified likelihood $p(d|\boldsymbol{\theta})$ and prior $p(\boldsymbol{\theta})$.

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

$$p(\boldsymbol{\theta}|d) \propto \frac{p(d|\boldsymbol{\theta})p(\boldsymbol{\theta})}{q_\phi(\boldsymbol{\theta}|d)}. \tag{3.1}$$

Given a set of N samples $\boldsymbol{\theta}ii = 1^N$ drawn from the approximate posterior distribution $q_\phi(\boldsymbol{\theta}|d)$,

$$w_i = \frac{p(d|\boldsymbol{\theta}_i)\, p(\boldsymbol{\theta}_i)}{q_\phi(\boldsymbol{\theta}_i|d)}, \qquad \tilde{w}_i = \frac{w_i}{\sum_{j=1}^{N} w_j}, \tag{3.2}$$

where $\tilde{w}_i$ are the normalized weights satisfying $\sum_i \tilde{w}_i = 1$. The expectation value of any function $f(\boldsymbol{\theta})$ with respect to the target posterior can then be estimated as

$$\mathbb{E}_p[f(\boldsymbol{\theta})] \approx \sum_{i=1}^{N} \tilde{w}_i\, f(\boldsymbol{\theta}_i). \tag{3.3}$$

In practice, the importance weights are computed using the likelihood function provided by `bilby`, which guarantees consistency with the detector noise model and with the specific data segment employed for parameter estimation. This procedure allows the neural posterior to accommodate moderate discrepancies

between the simulation-based training distribution and the true data likelihood without retraining the network. The quality of this reweighting step is characterized by the sampling efficiency, which quantifies how well the weighted samples approximate the target posterior distribution,

$$\epsilon = \frac{1}{N} \frac{(\sum_i \tilde{w}_i)^2}{\sum_i \tilde{w}_i^2}, \tag{3.4}$$

which characterizes the effective number of posterior samples that substantially contribute to the weighted estimate. In our applications, the efficiency ε generally remains at or above a few percent, showing that the learned posterior is sufficiently close to the true distribution to permit reliable importance reweighting.

3.2.6 Lensing Bayes Factor

To evaluate how strongly the data support the lensed hypothesis over the nonlensed hypothesis, the analysis considers the Bayes factor, defined as the ratio of the Bayesian evidences (marginal likelihoods) for the two competing models.

$$\log_{10} \mathcal{B}_{\text{lens}} = \frac{p(d|\mathcal{H}_{\text{L}})}{p(d|\mathcal{H}_{\text{NL}})}, \tag{3.5}$$

The Bayesian evidence associated with each hypothesis is computed via importance sampling. In this scheme, samples are drawn from the neural posterior estimator and then reweighted by the ratio of the true likelihood to the approximate network likelihood, following the procedure of [61]. This approach permits an efficient and unbiased estimate of the marginal likelihood, without the need to perform an explicit integration over the entire parameter space. Separate lensed and nonlensed models are trained to perform posterior inference for each hypothesis, using the training setup and validation methodology outlined in the preceding sections. The validation of the nonlensed network is presented in Chapter 6.1

3.3 Results

This section presents the verification of `DINGO-lensing` through an evaluation of its training performance and network validation, followed by applications to the detectability of lensed GWs in inference and an in-depth analysis of GW231123 using `DINGO-lensing`.

3.3.1 Verification of `DINGO-lensing`

In this section, the training results of the lensed network are presented and its performance is evaluated through a series of validation tests. Representative corner plots are also compared against `bilby` posteriors, demonstrating that `DINGO-lensing` accurately recovers point lens injections.

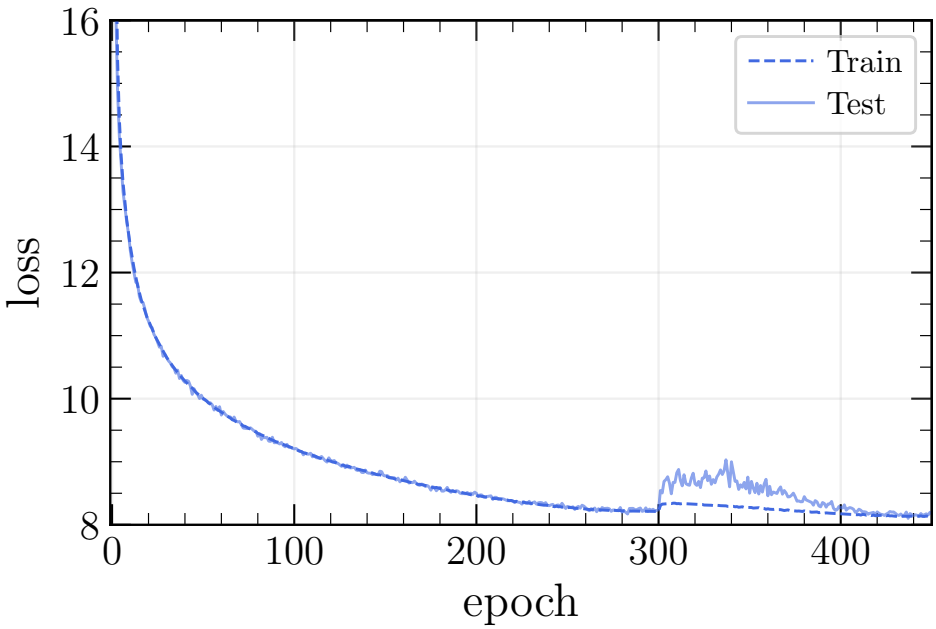

Figure 3.1: The loss as a function of the number of training epochs of the lensed networks, shown for the training set as dashed curves and for the testing set as solid curves.

3.3.1.1 Training performance

Figure 3.1 shows how the loss evolves as a function of training epoch for the model using O5 noises (left) and targeted model for GW231123 (right). The training loss (dashed) and test loss (solid) remain closely aligned throughout, indicating no signs of overfitting. The smoothness of both curves reflect stable optimization behavior. The loss curve plateaus after roughly 300 epochs suggesting that the network is no longer extracting additional information from the training set. The noticeable jump around epoch 300 coincides with the start of Stage 1 in the left panel, when extra layers are unfrozen. Training is terminated at epoch 450, as no further improvement is observed. The entire training run takes about 10 days on a single NVIDIA A100 PCIe GPU.

3.3.1.2 Network validation

Inference is carried out on 10^3 lensed injections generated from the trained model, drawing 10^4 posterior samples per injection from the lensed network. The reference phase is then reconstructed following [61]. Figure 3.2 presents the corresponding probability–probability plot, built directly from the posteriors produced by the lensed network without applying importance sampling. The diagonal line indicates perfect calibration, where x percent of true parameters lie within the x percent credible intervals of the inferred posteriors. The colored curves for individual parameters closely trace this diagonal, showing that the network produces well calibrated posterior distributions. This calibration demonstrates that the sampling scheme behaves as intended, a conclusion further supported by the Kolmogorov–Smirnov test p values listed in the legend [194]. The modest deviations seen for the lensing parameters Δt and $\mu_{\rm rel}$ remain within the expected three sigma statistical fluctuations, indicating robust performance even in the

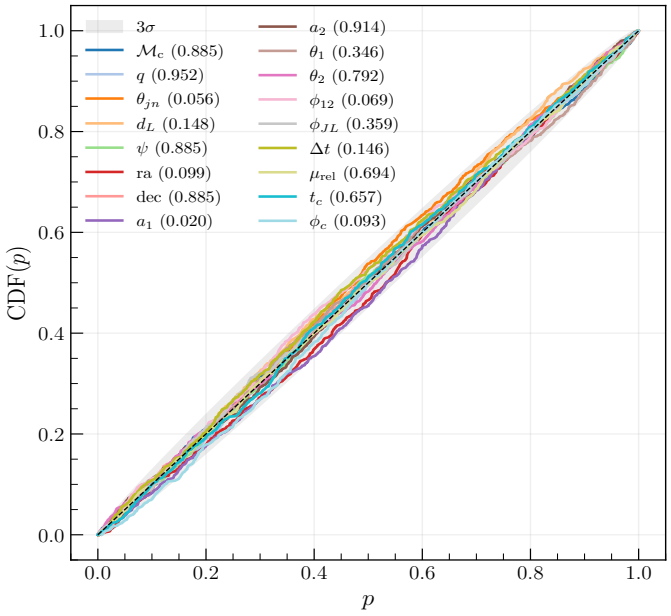

Figure 3.2: Probability–probability plot for the lensed neural posterior estimation models, constructed from 10^3 lensed injections without applying importance sampling. For each injection, a posterior is drawn from the model and the percentile rank of the injected parameter is computed. Each colored curve represents the cumulative distribution function of a given parameter, as indicated in the legend. The corresponding Kolmogorov–Smirnov test p values are also listed in the legend.

presence of lensing.

Figure 3.3 illustrates the distribution of sampling efficiencies ϵ after importance sampling is also analyzed. These efficiencies cover a wide range from roughly 10^{-4} to 0.1, with no apparent dependence on either the binary or lensing parameters. The effect of sampling efficiency on the Bayes factor is additionally investigated, revealing good consistency between `DINGO-lensing` and `bilby` for $\epsilon > 10^{-4}$ (See Appendix 6.2).

3.3.1.3 Comparison with traditional parameter estimation methods

Figure 3.4 displays a corner plot comparing posterior distributions inferred with `DINGO-lensing`, shown as solid blue curves, to those obtained with the reference `bilby` pipeline, shown as dashed gray curves, for a representative lensed injection with a network SNR of 16.18 at O5 sensitivity. The source parameters and priors used in the `bilby` run are summarized in Table 3.1.

The one and two dimensional marginalized posteriors agree closely between `DINGO-lensing` and `bilby` for both intrinsic parameters, such as $\mathcal{M}_c$, q, and $a_{1,2}$, and extrinsic parameters, such as d_L, θ_{JN}, and ψ. This shows that the neural posterior estimator reproduces the results of full Bayesian inference while requiring substantially less computation. The injected values, marked by solid gray lines, lie within the 68% credible regions (dark blue) and 95% regions (light blue), and the sampling efficiency is 2.3%. The Bayes factors obtained with `bilby` and `DINGO-lensing` are $\log_{10} \mathcal{B}_{\rm lens} = 19.0$ and 18.9, respectively, consistent within statistical uncertainties, confirming that `DINGO-lensing` can robustly evaluate the

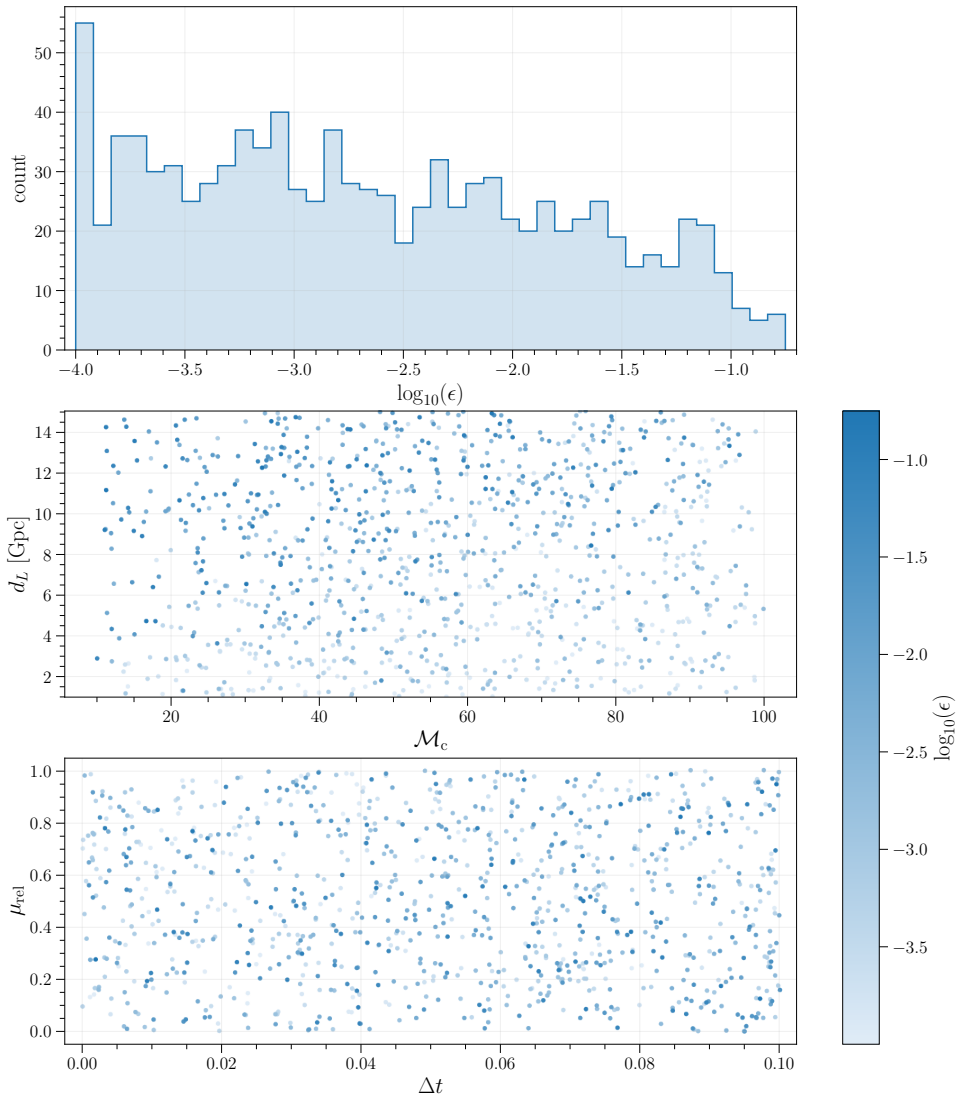

Figure 3.3: Distribution of the sampling efficiencies ϵ of 10^3 lensed injections for the lensed network. 10^5 samples are drawn for each injections The top panel shows the histogram of $\log_{10}(\epsilon)$, with most samples distributed between 10^{-4} and 10^{-1}. The middle panel illustrates the dependence of ϵ on the source parameters, where each point represents an injection colored by $\log_{10}(\epsilon)$ in the $(\mathcal{M}_c, d_L)$ plane. A weak trend is observed in which higher $\mathcal{M}_c$ and larger d_L correspond to an improved sampling efficiency. The bottom panel shows the variation of ϵ with the lensing parameters $(\Delta t, \mu_{rel})$, indicating no strong correlation but a broad spread consistent with the added complexity of the lensing parameter space.

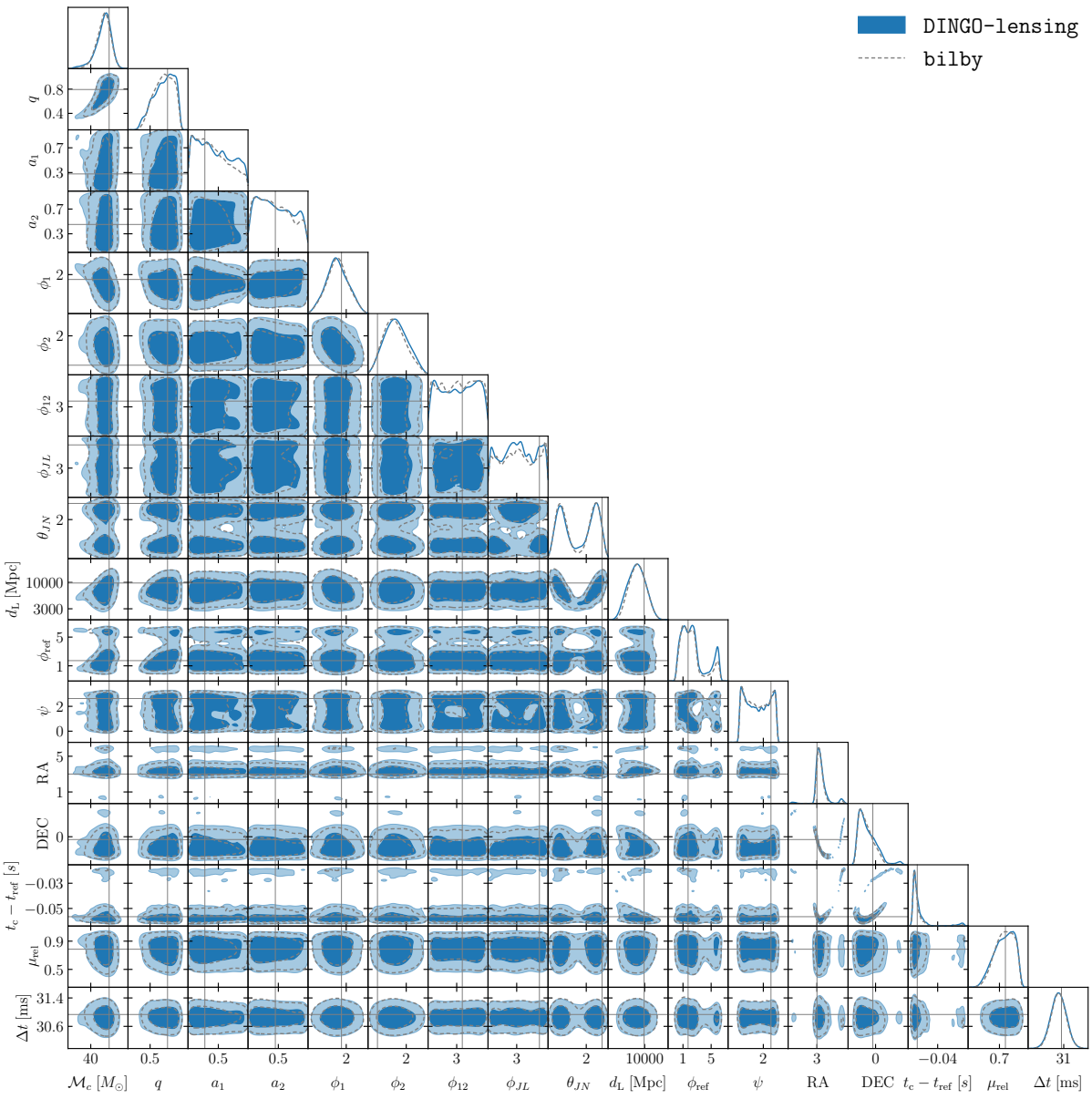

Figure 3.4: Posterior distributions for a representative injection inferred with `DINGO-lensing` using the O5 model, shown as solid blue curves, and with `bilby`, shown as dashed gray curves. The two methods provide mutually consistent posteriors and both well recover the injected parameter values, indicated by the solid gray lines and listed in Table 3.1.

lensing hypothesis while reducing the total inference time to about 40 seconds.

3.3.1.4 Recovery of GWs diffracted by point lenses

To assess how accurately the geometrical optics approximation captures lensing induced distortions in GW waveforms, inference is performed on a lensed signal produced by a point-mass lens. In this setup, the amplification factor has a closed form expression that depends on the redshifted lens mass M_{Lz} and the impact parameter y [26]. A point-mass lens invariably generates two images and also gives rise to diffraction effects, which are not explicitly represented in the parametrization of interfering chirps in stationary phase approximation and overlaping chirps.

In practice, the lensed waveform is generated using the point-mass lens implementation in `modwaveforms`, which follows the formulation of Reference [92]. Recall in Section 1.4.1.5, the lens mass and impact parameter map uniquely onto the effective lensing observables in geometrical optics limit, the relative time delay and relative magnification between the two images, allowing a direct comparison between the full point lens model and the simplified two image description.

$$\Delta t = \frac{4GM_{\mathrm{Lz}}}{c^3}\Delta T(y)\,, \tag{3.6}$$

$$\mu_{\mathrm{rel}} = \left|\frac{y\sqrt{y^2+4}-(y^2+2)}{y\sqrt{y^2+4}+(y^2+2)}\right|\,, \tag{3.7}$$

where the dimensionless time delay is determined by the impact parameter

$$\Delta T(y) = \frac{1}{2}y\sqrt{y^2+4} - \ln\left|\frac{y-\sqrt{y^2+4}}{y-\sqrt{y^2+4}}\right|\,. \tag{3.8}$$

The time delay scales linearly with the lens mass. For a lens with $M_{\mathrm{Lz}} = 100M_\odot$, the characteristic delay is of order $4GM_{\mathrm{Lz}}/c^3 \approx 2,\mathrm{ms}$. In the limit $y \to 0$, the two images approach equal magnification ($\mu_{\mathrm{rel}} \to 1$) and the time delay tends to zero, $\Delta t \to 0$.

Figure 3.5 shows the probability-probability plot used to assess the calibration of posterior inferences for the lensing parameters Δt and μ_{rel}, constructed from 10^3 simulated GW signals lensed by a point-mass, with 10^4 samples drawn from the lensed network for each injection. The simulated events are generated by sampling lens masses uniformly between $500M_\odot$ and $1500M_\odot$ and impact parameters uniformly between 0 and 1.5. This choice guarantees that the resulting time delays remain below 0.1, s, matching the maximum delay represented in the training set. Because the signal durations always exceed 0.1, s, these injections probe the wave-optics regime [66, 26].

Each curve displays the cumulative fraction of true parameter values contained within the nominal credible intervals, compared against the ideal uniform expectation shown as a black dashed line. The light gray band indicates the expected 3σ statistical variation for a perfectly calibrated inference under the null hypothesis. Although the curves for both Δt and μ_{rel} exhibit mild excursions beyond the 3σ region, these deviations are small and do not point to significant systematic biases. The corresponding Kolmogorov–Smirnov test p values are listed in the legend. Taken together, these results show that, even when recovering point-mass–lensed signals within the geometrical optics approximation, the inferred

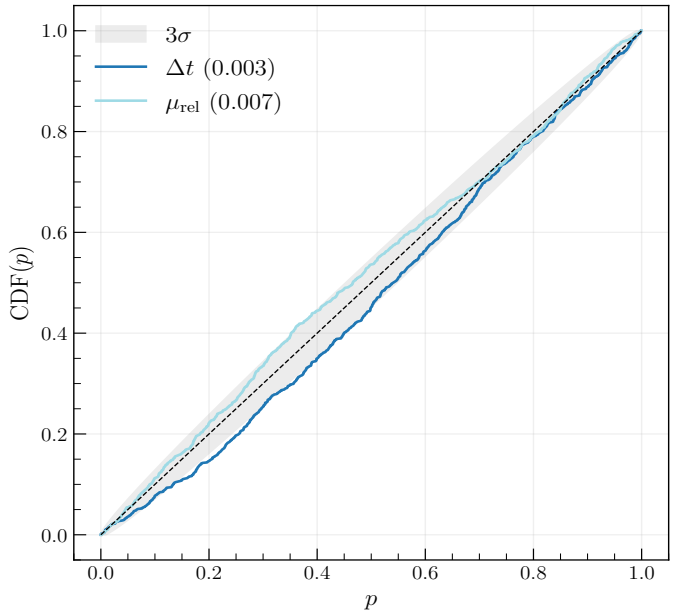

Figure 3.5: Probability–probability plot showing the calibration of the posterior distributions for the lensing parameters Δt and $\mu_{\rm rel}$ using the O5 model, obtained from 10^3 simulated point-mass–lensed GW injections. Each curve represents the cumulative fraction of true parameter values contained within a given credible level, compared against the ideal uniform expectation indicated by the black dashed line. The light gray band marks the expected 3σ statistical fluctuations for a perfectly calibrated inference. The analysis recovers signals generated with a point lens model in the geometrical optics regime and demonstrates accurate reconstruction of both the time delay and relative magnification distributions.

posteriors remain well calibrated and any residual systematics are negligible compared to statistical fluctuations.

To further illustrate the quality of individual recoveries, we highlight a representative lensed signal with a duration of roughly 0.1s. Figure 3.6 shows a subset of the full seventeen dimensional corner plot, demonstrating the recovery of a lensed GW signal within the `DINGO-lensing` framework for a point-mass lens with $M_{\rm Lz} = 900 M_\odot$ and $y = 0.3$. A total of 10^5 posterior samples are obtained from the lensed network using importance sampling.

The two dimensional posterior distributions place tight constraints on the lensing parameters, indicating that the model successfully captures the distortions induced by the lens. The inferred values $\Delta t = 10.47 \pm 0.57$ms and $\mu_{\rm rel} = 0.6 \pm 0.17$ are in good agreement with the geometrical optics expectations of $\Delta t = 11$ms and $\mu_{\rm rel} = 0.5$. This consistency shows that the geometrical optics approximation can accurately reproduce the effective lensing signatures encoded in the waveform, even when the signal lies in the wave-optics regime.

This demonstration does not rely on a fully self consistent point lensed source model for parameter estimation. Although such models are analytically tractable, they omit finite source effects and frequency-dependent interference, which become important outside the strict geometrical optics limit. Instead, the aim here is to validate the ability of the `DINGO-lensing` network to recover the effective observables Δt and $\mu_{\rm rel}$ that characterize lensing distortions within this approximation. A direct comparison with a

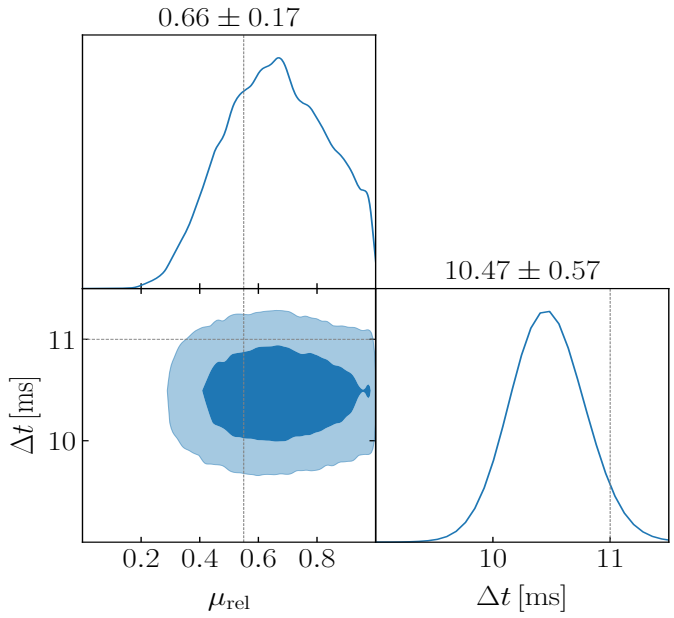

Figure 3.6: The injection corresponds to a point-mass lens with $M_{\mathrm{Lz}} = 900 M_{\odot}$ and $y = 0.3$. The inferred values by the O5 lensed model, $\Delta t = 10.47 \pm 0.57$ms and $\mu_{\mathrm{rel}} = 0.66 \pm 0.17$, are consistent with the geometrical optics predictions of $\Delta t = 11$ ms and $\mu_{\mathrm{rel}} = 0.5$. This agreement demonstrates that the geometrical optics approximation can accurately reproduce lensing induced waveform distortions.

full Bayesian recovery using `bilby` is not performed, both because of the substantial computational cost and because such analyses have already been presented elsewhere [136, 195]. This test therefore offers a controlled and efficient validation that the network reproduces geometrical optics predictions with high fidelity.

3.3.1.5 Detectability of lensed gravitational waves in inference

`DINGO-lensing` is used to quantify how detectable gravitational waveform distortions from lensing are in a fiducial Hanford–Livingston O5 detector configuration. A large ensemble of lensed and nonlensed signals is simulated by drawing sources uniformly in comoving volume between 1 Gpc and 15 Gpc, with chirp masses between 30 and 60 $M_{\odot}$, mass ratios $q > 0.3$, and all remaining parameters sampled from the same fiducial distributions adopted for network training. The reference trigger GPS time is fixed to 100 s with a uniform jitter of ±0.1 s, and only events with network SNR greater than 8 are retained. For each sampled parameter set, a frame file is produced and detector strain data are generated by injecting the signal into Gaussian noise consistent with the assumed noise power spectral density, after which each frame is analyzed by both the lensed and nonlensed networks to compute the lensing Bayes factor via importance sampling. This simulation pipeline is validated against `bilby` for events covering a range of sampling efficiencies, as described in Appendix 6.2, and the Bayes factors from `DINGO-lensing` and `bilby` are found to agree for $\epsilon > 10^{-4}$, a value adopted as a conservative lower efficiency threshold in the analysis.

Lensing candidates are identified using both the lensing parameters themselves and summary statistics

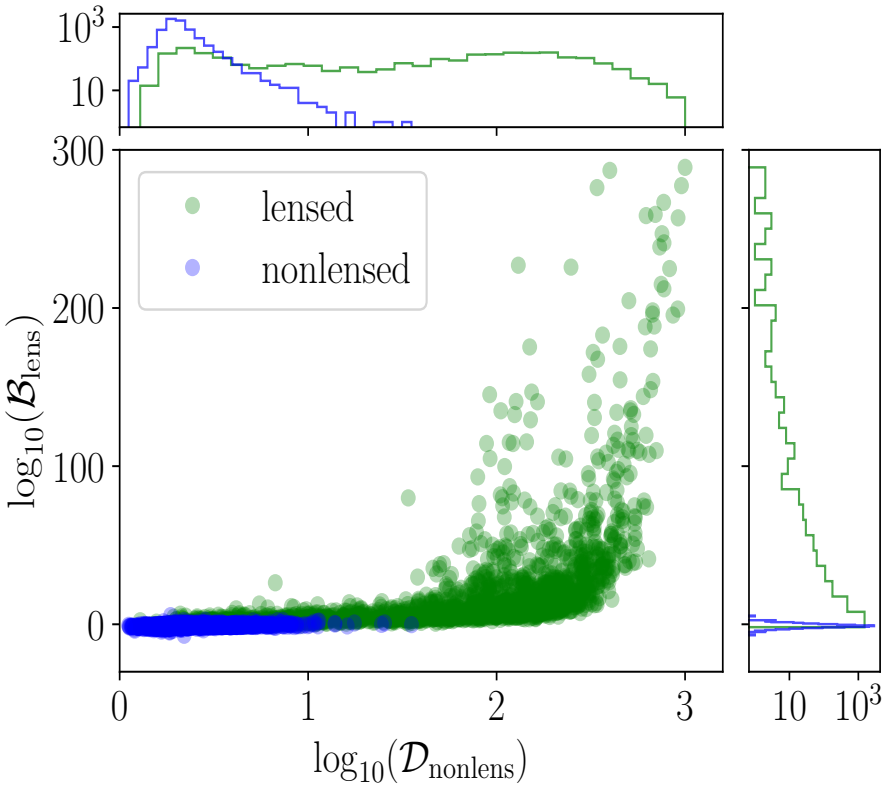

Figure 3.7: Distribution of the lensing Bayes factors $\log_{10}\mathcal{B}_{\rm lens}$ and the distance from the nonlensing hypothesis $\mathcal{D}_{\rm nonlens}$ for the set of 2,561 lensed and 10,858 nonlensed simulated events analyzed using the O5 network.

derived from their posteriors. A direct approach is to inspect the posterior samples in the $(\Delta t, \mu_{\rm rel})$ plane, where significant deviations away from $(0, 0)$ already signal tension with the nonlensing hypothesis. To quantify this, a Gaussian distance from the point $(0, 0)$ is computed following Reference [196], defining the distance away from the nonlensing hypothesis, $\mathcal{D}_{\rm nonlens}$, normalized by the posterior uncertainties so that well constrained, clearly offset posteriors yield larger $\mathcal{D}_{\rm nonlens}$ values and typically satisfy $\mathcal{D}_{\rm nonlens} \gtrsim 1$ because both parameters are bounded below by zero.

Figure 3.7 compares the distributions of $\mathcal{D}_{\rm nonlens}$ and the lensing Bayes factor $\mathcal{B}_{\rm lens}$ for 2,561 lensed and 10,858 nonlensed simulations, showing that the two statistics are strongly correlated, with lensed simulations producing a pronounced high–$\mathcal{D}_{\rm nonlens}$, high–$\mathcal{B}_{\rm lens}$ tail, while nonlensed simulations occupy a much narrower range in both quantities. The Bayes factor distributions are examined in more detail in Fig. 3.12 using complementary cumulative distribution functions, where lensed simulations preferentially yield large Bayes factors, with 53% exceeding 5, but the significance of any single candidate must still be evaluated against a background distribution under the nonlensing hypothesis, motivating the calculation of a false alarm probability, as done in previous `DINGO` background studies (e.g., for eccentric signals).

Within the simulated background, only 2% nonlensed events favor lensing with $\mathcal{B}_{\rm lens} > 0$, but the background distribution depends sensitively on the assumed source population. As illustrated in Fig. 3.12, nonlensed systems with higher chirp masses of [30,40]$M_\odot$ are more likely to mimic lensing signatures

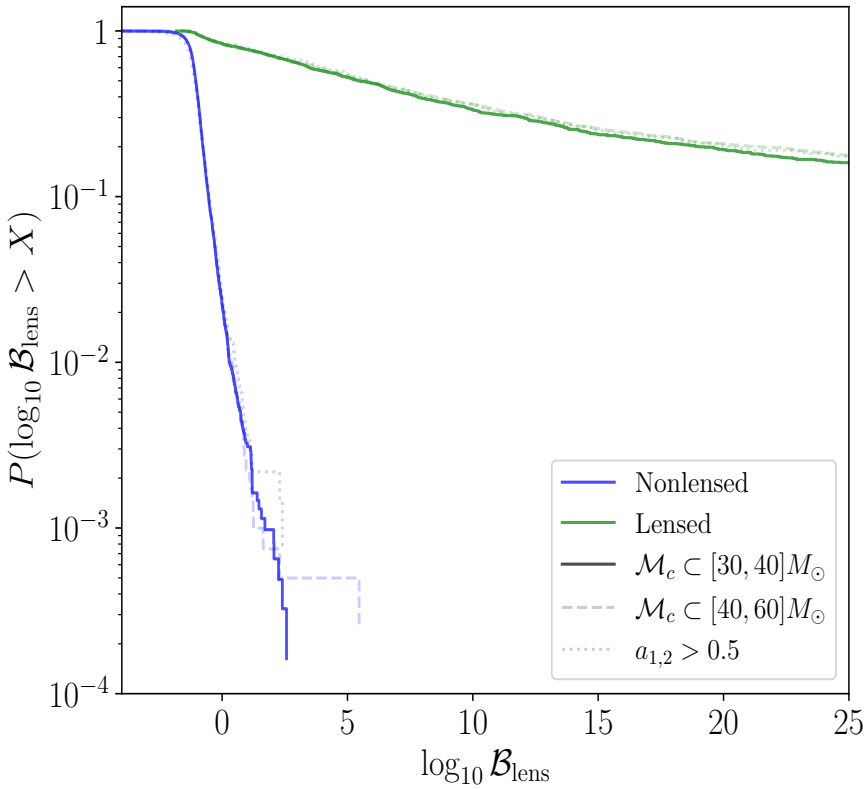

Figure 3.8: Complementary cumulative distribution function of lensing Bayes factors $\mathcal{B}$lens for simulated lensed and nonlensed events at O5 sensitivity. Solid curves correspond to events with chirp masses in the range [30,40]$M_\odot$, while dashed curves correspond to [40,60]$M_\odot$. For each chirp mass range, dotted curves indicate the subset of events with both component dimensionless spin magnitudes smaller than 0.5.

than those of [40,60]$M_\odot$. In addition, restricting to nonlensed simulations in which either component has a dimensionless spin magnitude above 0.5 further increases the false alarm rate, consistent with the expectation that highly spinning systems with stronger precession and shorter signal durations are more prone to resemble lensing induced waveform distortions by chance.

3.3.2 GW231123 analysis with `DINGO-lensing`

In this section, we perform a detailed analysis to GW231123, an event with the highest support to GWTC-4.0 by assessing the statistical significance of lensing. Training and validation of lensed and nonlensed GW231123 network are discussed in Section 6.4 and 6.5 respectively. We further discuss the strategiesstrategies of utilizing `DINGO-lensing` for future lensing discovery.

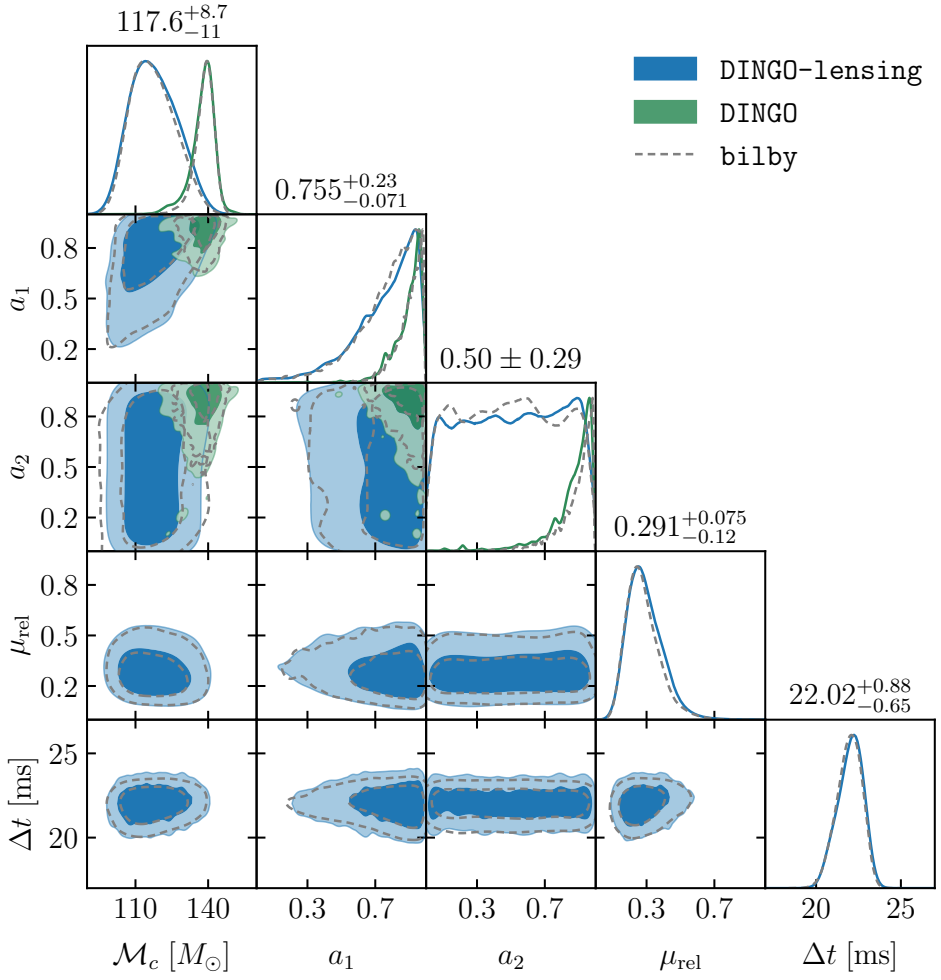

Figure 3.9: Source characterization of GW231123 under the lensed hypothesis using `DINGO-lensing` (blue) and under the nonlensed hypothesis using `DINGO` (green). The posterior distributions obtained with `DINGO-lensing` and `DINGO` are consistent with those recovered using `bilby`. All quoted credible intervals correspond to the `DINGO-lensing` inference.

3.3.2.1 Statistical signficance of lensing

To quantify the statistical significance of the apparent preference for lensing in GW231123, a population of GW231123-like events is simulated with component masses chosen such that the chirp mass lies between 90 and 160, $M_\odot$, mass ratios between 0.2 and 1, and luminosity distances distributed uniformly in comoving volume from 0.6 to 8, Gpc, within the sensitive range of the current detector network. Spin magnitudes are drawn up to 0.99 with isotropic orientations, and only events with Hanford–Livingston network signal to noise ratio greater than 8 are retained. Within this simulated background, many nonlensed events nonetheless produce lensing posteriors whose support extends away from zero, as illustrated for a representative subset in Fig. 3.10, reflecting the self similarity of short duration, high mass signals that can mimic lensing distortions.

The distance of the lensing posteriors from zero in the $(\Delta t, \mu_{\rm rel})$ plane serves as a preliminary indicator of lensing and can be evaluated extremely quickly using neural posterior estimation, typically within

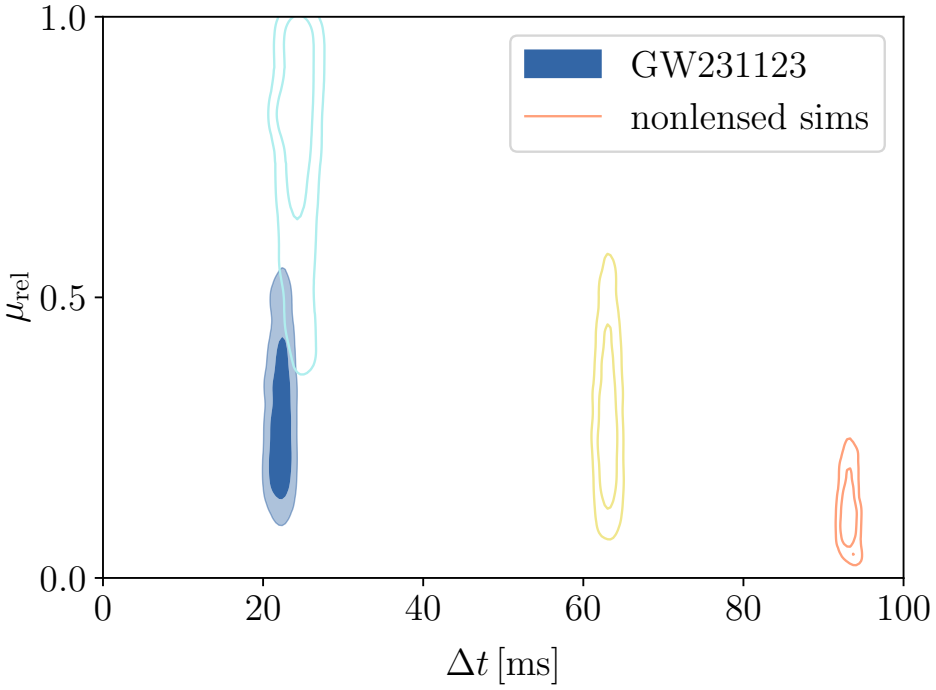

Figure 3.10: Inference for the lensing parameters. The filled contours show the results for GW231123, while the unfilled contours correspond to nonlensed simulations of comparably massive binaries.

a few seconds per event, enabling large ensembles of GW231123-like simulations to be analyzed for significance studies.

In many simulations the recovered Δt shows a multimodal structure, with peaks separated by roughly the characteristic period of the waveform near merger. This behavior reflects the self similarity of very short signals like GW231123, where only a few cycles are observable and time shifts by one period can produce waveforms that remain highly correlated. Figure 3.11 illustrates this effect for GW231123 by comparing the whitened best fit lensed waveform, the corresponding unlensed waveform obtained by dividing out the lensing amplification, and a nonlensed waveform fitted under the no lensing hypothesis, all whitened using the Hanford noise power spectral density for the event. The inferred time delay between the two chirps, $\Delta t \approx 22,\mathrm{ms}$, is consistent with the instantaneous waveform period near merger at the reference time $t_{\rm ref}$, highlighting the degeneracy between lensing induced delays and periodic self similarity in such short signals.

The Bayes factor for GW231123 is evaluated against a background of 70,000 GW231123 like nonlensed simulations, whose complementary cumulative distribution function is shown in Fig. 3.12. In this background, 8% of nonlensed events favor lensing with $\mathcal{B}_{\rm lens} > 0$, and the distribution exhibits a long tail toward large Bayes factor values, placing GW231123 at a false-alarm probability corresponding to 4σ under the nonlensing hypothesis.

This is contrasted with a set of 1,000 GW231123 like lensed simulations, within which 58% of events produce a Bayes factor larger than that measured for GW231123. Even though short, very massive mergers like GW231123 make it intrinsically challenging to reach high detection significance, the same

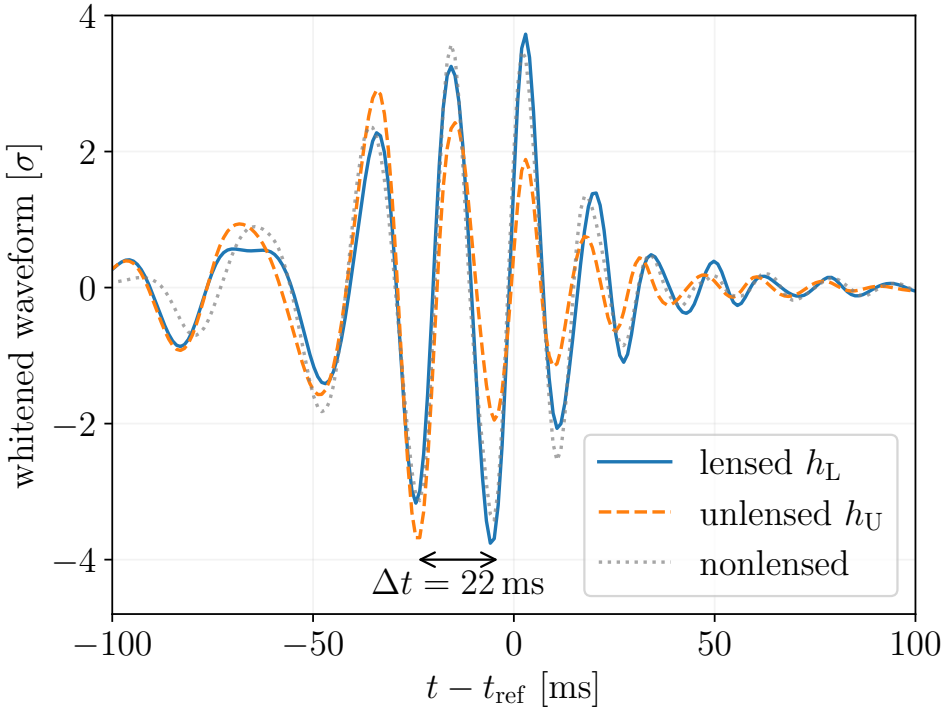

Figure 3.11: Whitened best fit lensed (solid), unlensed (dashed), and nonlensed (dotted) waveforms for GW231123. The inferred time delay between the two chirps, Δt, is consistent with the instantaneous waveform period near the reference time.

population study indicates that about 40% of genuinely lensed simulations would favor lensing with a significance above 5σ, demonstrating that such strong detections are achievable in principle for similar events.

3.3.2.2 Detection strategies of lensing using deep learning

The analysis of GW231123 serves as a testbed for designing detection strategies for lensed gravitational waves. A two step workflow is envisaged, beginning with a fast identification stage in which networks are trained on average detector configurations and representative source populations, providing a flexible machine learning model that can rapidly flag promising lensing candidates for follow up.

Within this first stage, `DINGO-lensing` can identify candidates either by checking whether the lensing posteriors in $(\Delta t, \mu_{\text{rel}})$ are clearly displaced from zero, as in Fig. 3.10, or by computing the lensing Bayes factor and comparing it to simulated background distributions, as in Fig. 3.12.

A second, focused stage involves training dedicated networks on the most significant candidates, as demonstrated here for GW231123, to capture event specific features such as detector noise nonstationarity, waveform systematics, network configuration, and data quality issues. Multiple such focused networks can be trained under different modeling assumptions, providing tailored background distributions needed to assess the statistical significance of individual lensed candidates, while retaining the orders of magnitude computational speedup of `DINGO-lensing` over traditional parameter estimation for both training and

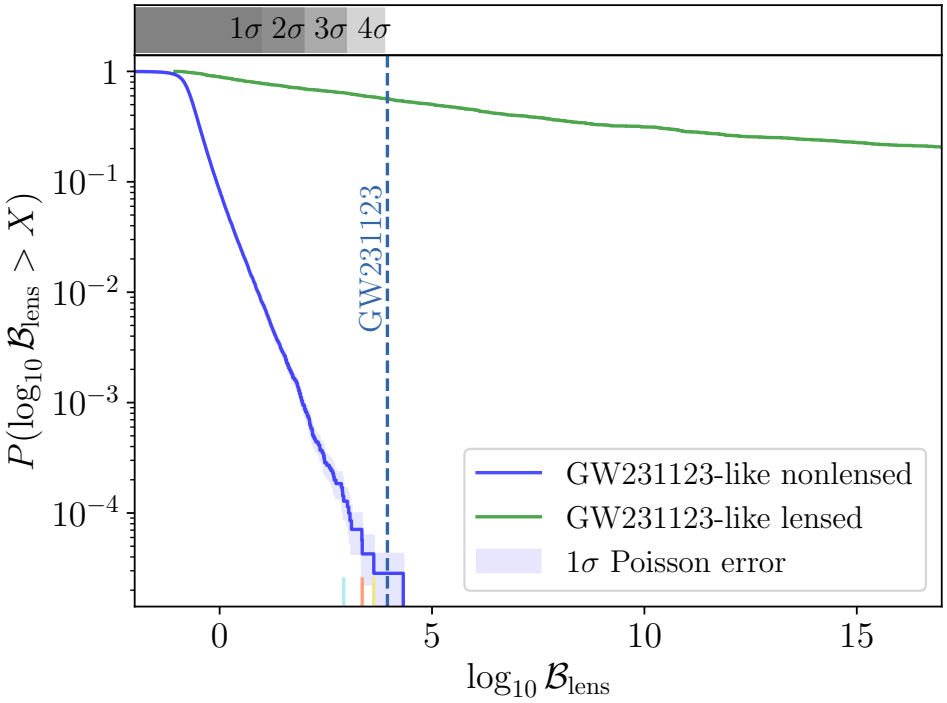

Figure 3.12: Complementary cumulative distribution function of the lensing Bayes factors $\mathcal{B}_{\rm lens}$ for simulated GW231123 like events. The lensing Bayes factor of GW231123, xxx, is marked by the vertical dashed line. The shaded band represents the Poisson uncertainty given by $1/\sqrt{N}$. Short vertical markers denote the Bayes factors of the representative examples shown in Fig. 3.10.

inference.

The statistical significance of lensing induced waveform distortions can be even higher in scenarios where strong lensing produces multiple observable images of the same source, particularly when some images show additional distortions that are difficult to mimic with alternative effects. Such configurations are expected for populations of microlenses [73, 47] or compact dark matter subhalos located near critical curves [117], where interference and diffraction signatures in GWs provide sensitive probes of small scale structure.

3.4 Conclusions

In this chapter, we present `DINGO-lensing`, a simulation-based inference framework that enables rapid and precise parameter estimation of gravitationally lensed GW signals through deep learning. By directly learning the posterior distribution from simulated datasets, `DINGO-lensing` bypasses the severe computational demands of conventional Bayesian sampling, while maintaining full statistical interpretability.

3.4.1 Inference of distorted gravitational wave with `DINGO-lensing`

We train neural posterior estimators on a dataset of 5×10^6 simulated waveforms, building separate models for lensed and nonlensed signals that yield well-calibrated posterior distributions. Their performance is validated through probability–probability plots and direct comparisons with standard Bayesian parameter estimation using `bilby` [52]. Across intrinsic, extrinsic, and lensing parameters, the models show remarkable consistency with traditional methods, achieving speedups from weeks to mere tens of seconds per event.

Networks trained at LIGO design sensitivity demonstrate the framework's capabilities. We show that lensing parameters are recovered with high fidelity, with typical uncertainties in lensing time delays on the order of milliseconds. In our fiducial two-image lens model featuring overlapping chirps, the method successfully reconstructs gravitational waves diffracted by point-mass lenses. To evaluate event detectability, we generate and analyze thousands of simulated lensed and unlensed signals, estimating their inferred lensing Bayes factors. For representative signals, the resulting Bayes factor distributions are clearly separable, highlighting strong discriminative power.

However, we find that the significance of a potential detection depends sensitively on the reference background population. In particular, when comparing to populations of heavier and more rapidly spinning binaries, higher Bayes factors are needed to achieve the same false-alarm probability as for populations dominated by lighter, slowly spinning binaries. A comprehensive detectability forecast incorporating realistic population models will be addressed in future work.

The detection of lensed GWs demands scalable, statistically robust methods, and `DINGO-lensing` contributes at two complementary stages. First, it facilitates rapid candidate identification using networks trained on average detector performance for a given observing run. Such models should cover a wide range of binary masses, ensuring compatibility with diverse waveform durations and lensing time delays short enough to produce a single distorted signal. Candidate selection can proceed either by (1) comparing the inferred lensing posterior to the unlensed hypothesis ($\Delta t = \mu_{\text{rel}} = 0$) or (2) computing the Bayes factor via importance sampling between lensed and unlensed hypotheses, benchmarked against precomputed background simulations. These background sets should incorporate both uniform priors and astrophysically motivated populations that include realistic source distributions and lensing optical depths to avoid bias.

Second, `DINGO-lensing` supports the training of event-specific neural networks for high-significance candidates. Tailored models are crucial to account for waveform systematics, and individual source characteristics, which we will investigate in Section 3.3.2. Such dedicated training enables the accurate estimation of background distributions and statistical significance at discovery-level confidence, which may require between thousands and millions of simulations to achieve $3\sigma/5\sigma$ confidence levels.

In summary, `DINGO-lensing` offers a scalable, efficient analysis strategy for GW lensing in upcoming LVK observing runs. Its computational speed makes it feasible to perform extensive background estimation and population-level studies that would otherwise be computationally prohibitive. Future developments include extending the framework to more complex lens models, incorporating waveform systematics, and embedding `DINGO-lensing` into real-time end-to-end pipelines for the identification and characterization

of lensed events. Because real detector noise is non-Gaussian and nonstationary, these factors may affect neural posterior fidelity and lensing detectability. Although their quantitative impact remains to be explored, idealized noise assumptions likely overestimate sensitivity. Upcoming work will systematically investigate the robustness of neural posterior estimation and detection performance under various glitch mitigation tools, such as `BayesWave` [197, 198] and `gwsubtract` [199, 200].

3.4.2 A deep dive into GW231123

We apply `DINGO-lensing` to a detailed reanalysis of GW231123, presently the highest ranked lensed candidate. In this case, the inferred false-alarm probability associated with the apparent preference for lensing is approximately 4σ and can plausibly be attributed to the self similarity of the signal. Moreover, many simulated GW231123 like lensed signals yield Bayes factors well above the 5σ threshold, making the discovery of lensing through waveform distortion a promising observational avenue. Additional sources of spurious lensing evidence include waveform systematics and nonstationary noise features [5]. These effects would only act to increase the false-alarm probability, so our estimate should be viewed as an upper bound on the statistical significance of a lensing interpretation for this event.

We show that `DINGO-lensing` enables efficient, systematic evaluation of the statistical significance of lensed candidates by analyzing large ensembles of lensed and nonlensed events, which is essential for establishing the first robust detection of lensed GWs. The same framework also supports rapid identification of promising candidates for electromagnetic follow up, which is particularly important for multimessenger lensing involving binary neutron stars [8]. Although our study focuses on gravitational lensing, the methodology is broadly applicable to searches for waveform modifications arising from other physical effects, such as departures from general relativity or environmental influences. Taken together, these results indicate that deep learning accelerated inference will play a central role in uncovering new phenomena in GW astronomy.

Chapter 4

TESLA-X: An effective method to search for sub-threshold lensed gravitational waves with a targeted population model

In GW searches, events with amplitudes high enough to be robustly distinguished from detector noise are labeled super-threshold, while weaker signals that cannot be reliably separated from noise are classified as sub-threshold. Lensing can produce multiple images of the same GW source, each with a different magnification factor $\sqrt{\mu_j}$ and arrival time (See Section 1.4.1.4). A lensed GW signal can fall below the standard detection threshold for two distinct reasons. First, in the wave-optics regime, frequency-dependent waveform distortions can degrade the match with nonlensed templates, thereby reducing the recovered SNR (See Chapter 2). Second, in the strong lensing regime, certain images can be demagnified, leading to a suppression of the waveform amplitude relative to the nonlensed signal. In this work, we focus on the latter scenario and investigate whether such demagnified, sub-threshold lensed counterparts can be identified when at least one associated image is detected as a super-threshold event.

Matched-filtering pipelines with dedicated adaptations have been used to search for sub-threshold strongly-lensed counterparts to confidently detected events from the LVK Collaboration's first three observing runs (O1, O2, and O3) [127, 20, 16]. These efforts include methods that specifically target additional, demagnified images corresponding to known catalog events. A central focus of this work is the Targeted sub-threshold Lensing Search (TESLA) method, which is built on the `GstLAL` matched-filtering pipeline and is explicitly optimized to recover potential sub-threshold lensed counterparts associated with identified super-threshold GW events.

4.1 Current efforts to search for strongly-lensed sub-threshold images

Two targeted search methods have been proposed and implemented to identify potential sub-threshold lensed counterparts associated with already-detected (super-threshold) events [18, 128]. One method is to utlize a single template bank using the maximum likelihood sample from the posterior [128]. This section focus on another approach that is `GstLAL`-based Targeted sub-threshold Lensing Search (TESLA) [18].

4.1.1 Overview of Targeted sub-threshold lensing search (TESLA) method

TESLA builds on the matched-filtering pipeline `GstLAL` and exploits this structure to search for sub-threshold, strongly-lensed counterparts to known super-threshold events. In a generic compact binary coalescence search, the goal is to cover the full intrinsic parameter space allowed by the detector bandwidth, which requires constructing a template bank with more than a million waveforms. This large bank yields many triggers, whether produced by noise or true signals, and the resulting increase in trials factor effectively raises the noise background, making it difficult to uncover weak signals such as demagnified lensed copies. TESLA takes advantage of the fact that, under strong lensing, multiple images from the same source share identical intrinsic parameters (component masses, spins, and hence intrinsic waveform), differing only in amplitude and arrival time. For each super-threshold target event, Bayesian parameter estimation is first used to obtain a posterior distribution over the source parameters, which defines a signal subspace consistent with the observed event [201, 52, 53, 135]. However, noise fluctuations and non Gaussian artifacts can cause weak counterparts to be recovered by templates lying well outside the high posterior density region, so the posterior alone is insufficient to define the search region [18].

To address this, TESLA explicitly incorporates both signal and noise behavior when constructing a targeted template bank. Posterior samples of the target event are ordered by decreasing log likelihood, and for each sample a set of injections is created: one with identical parameters to the sample and nine additional injections with increased luminosity distance, reducing the optimal SNR according to the scaling relation

$$D_L \propto \frac{1}{\rho_{\mathrm{opt}}}. \tag{4.1}$$

to mimic demagnification. These injections are required to have detector SNR at least four in at least one detector so that they can be registered as triggers without lowering the single detector threshold, which would otherwise sharply increase the noise trigger rate. The ensemble of injections represents plausible sub-threshold lensed counterparts. Each is added to real data and recovered using a full matched-filtering search with the original, large template bank. Templates that successfully recover these injections with false-alarm rate no worse than ≤ 1 in 30 days are retained and combined into a reduced, targeted template bank. This bank is then used to search the data for candidate sub-threshold lensed counterparts to the chosen super-threshold event.

The output of TESLA is a ranked list of candidate counterparts, ordered by increasing false-alarm rate. These ranking statistics do not directly give the probability that a candidate is truly a lensed counterpart but instead provide a priority ordering for subsequent follow up analyses such as detailed parameter estimation and lens modeling.

4.1.2 Limitation of the TESLA method

TESLA's performance requires resolving two key remaining issues that affect template bank optimality and parameter–space weighting.

In the original `GstLAL` searches, the full template bank is constructed over mass and spin ranges motivated by detectable compact binaries: binary neutron stars, neutron star–black hole systems, and binary black holes—and is placed densely enough that the maximum SNR loss from discrete template spacing is limited to about three percent relative to the optimal SNR, defined as the SNR obtained when the template exactly matches the signal parameters [23]. In traditional TESLA, a reduced template bank is built by discarding templates that do not recover any of the lensed injections associated with a given super-threshold target. Because this reduced bank is selected from a finite, limited set of injections, it no longer shares the formal optimality guarantees of the original bank. As a result, some strongly-lensed sub-threshold signals with optimal detector SNR close to the detection threshold, for instance $\rho_{\mathrm{opt}} \approx 4.12$, can fail to be recovered. Ensuring that the reduced TESLA bank retains near opimal coverage is therefore crucial for maintaining maximal sensitivity to potential strongly-lensed sub-threshold events.

Traditional TESLA also does not employ an explicit population model for sub-threshold counterparts (See the introduction of population model in Section 1.2.3.4). In practice, this means the analysis implicitly treats all templates in the reduced bank as equally likely to recover a lensed counterpart, regardless of how closely they match the intrinsic parameters of the known super-threshold event. Since the goal is to find sub-threshold images sharing the same intrinsic source properties as the target, introducing a population model peaked around the recovered masses and spins of the super-threshold signal can substantially increase sensitivity in the most relevant region of parameter space. Furthermore, this population model should be SNR dependent, reflecting the fact that weaker, sub-threshold signals are more strongly influenced by noise and can be recovered by templates that deviate from those favored for the loud, super-threshold detection.

4.2 Methods

To address the limitations of the traditional TESLA method, we first construct an optimally targeted template bank, and subsequently introduce a targeted population term into the ranking statistic of the matched-filtering search pipeline `GstLAL`. In this section, we introduce the essential concepts for the TESLA-X methods in Section 4.2.1 and detail the methods of the proposed approach in Section 4.2.3.

4.2.1 Optimal targeted template bank

The injection campaign's sole role is to identify which part of parameter space should be prioritized for strongly-lensed sub-threshold counterparts to a given target, but any such campaign is inherently approximate because it uses finitely many injections with discretized SNR values and thus cannot span the full continuum of possible sub-threshold signal strengths. Templates that succeed in recovering these injections should therefore be interpreted as mapping out the region in which real counterparts are likely

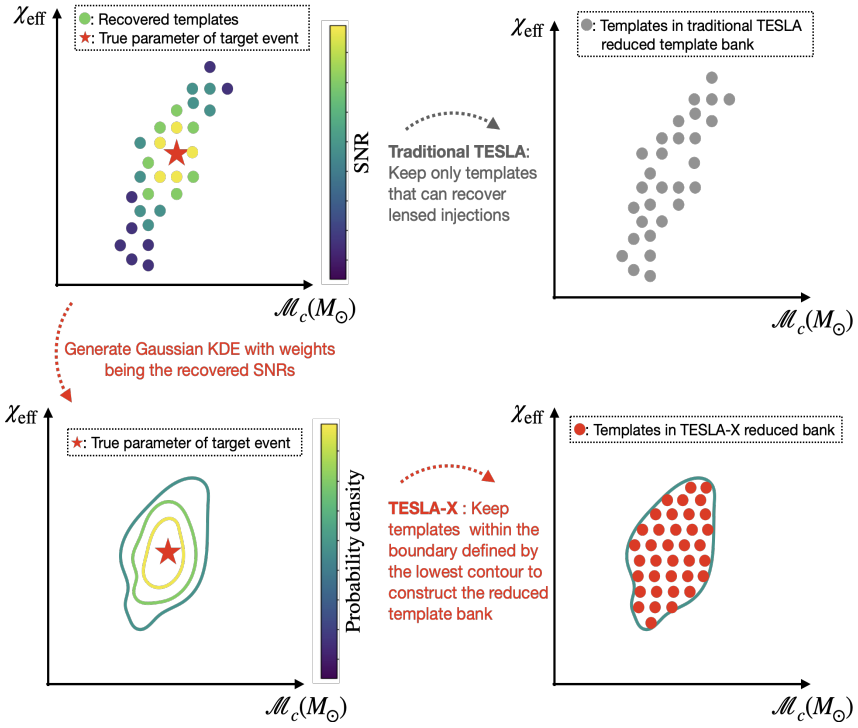

Figure 4.1: A schematic illustration of the construction of a reduced template bank using injection results, comparing the Traditional TESLA method shown in the top right panel with the proposed TESLA-X approach illustrated in the bottom left and right panels.

to be found, rather than as the only templates capable of detecting them.

On this basis, the injection campaign is used to define a boundary in parameter space (See Fig, 4.1), and all templates whose parameters fall within this boundary are retained in the reduced bank, regardless of whether they individually recovered an injection. Within this restricted region, the reduced bank then inherits the optimality properties of the original full bank, ensuring that any lensed counterpart inside the boundary is recovered with less than roughly three percent SNR loss, thereby preventing missed detections due to excessive mismatch.

4.2.2 Targeted population model for strongly-lensed sub-threshold signals

Under the strong lensing hypothesis, all images share the same intrinsic parameters as the super-threshold target, so the true counterpart population is expected to cluster near the signal subspace defined by the target's posterior, with probability tapering off away from that region.

Figure 4.2 illustrates the construction of the targeted population model. During the traditional TESLA injections, both the recovered SNRs and the identities of the recovering templates are recorded, revealing that weak sub-threshold signals, being more susceptible to noise, can be recovered by templates whose parameters differ significantly from those of the target, whereas templates closer to the target posterior tend to recover injections with higher SNR. This information is used to build a Gaussian kernel density estimate, KDE($\vec{\gamma}$), over the intrinsic source parameters $\vec{\gamma}$ (chosen to match the template bank parameters, e.g. component masses and spins), with each template weighted by the recovered SNR of the injections

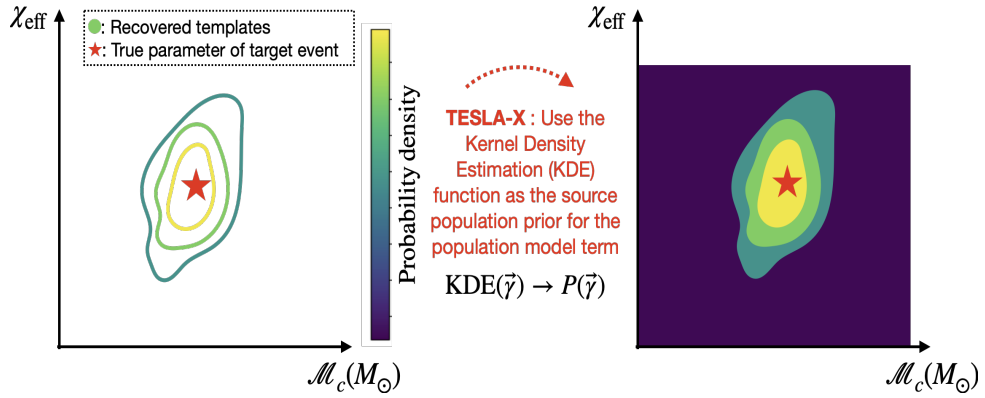

Figure 4.2: A schematic illustration showing how injection campaign results are used to build a targeted population model within the TESLA-X framework.

it recovers. The resulting KDE assigns larger probability density to regions near the targeted signal subspace, consistent with the expectation that true lensed counterparts should resemble the target's intrinsic parameters.

This KDE is then interpreted as the probability density function for the population of possible strongly-lensed sub-threshold counterparts and is used, following the method of [36], to compute $P(\vec{\theta} \mid \textbf{signal})$, the probability that a signal is recovered by a template with parameters $\vec{\theta}$. Crucially, this probability is made SNR dependent, so templates farther from the target subspace are not unduly penalized when they recover very weak sub-threshold signals, which are more easily displaced in parameter space by noise. Together, the robust reduced bank and the SNR-dependent KDE population model replace the implicit uniform prior of traditional TESLA with a physically motivated source model, thereby improving overall sensitivity to strongly-lensed GW counterparts.

4.2.3 Overview of the TESLA-X pipeline

The overall TESLA-X workflow is summarized in Fig. 4.3 and proceeds as follows. The search starts from a target super-threshold GW event found in a standard `GstLAL` run with the full template bank. Bayesian parameter estimation for this event provides posterior samples of the source parameters, which are then used to generate simulated lensed counterparts: for each posterior sample, one injection is created with the same optimal SNR and nine additional injections with smaller optimal SNRs obtained by increasing their effective luminosity distances, subject to the requirement that at least one detector has $\rho \geq 4$ so that the signal can be registered as a trigger. These injections, which represent possible sub-threshold lensed images of the target, are added to real detector data and recovered with `GstLAL` using the full template bank; up to this point, the procedure coincides with traditional TESLA.

After the injection campaign, the templates that successfully recover the simulated injections are collected along with their observed SNRs and can be plotted in, for example, the $(\mathcal{M}_c, \chi_{\text{eff}})$ plane, with color indicating the recovered network SNR. T raditional TESLA would retain only these "hit" templates to form its reduced bank, leaving holes in parameter space. TESLA-X instead constructs a Gaussian kernel density estimate $f(\vec{\gamma})$ over the template parameters $\vec{\gamma}$ (such as $\mathcal{M}_c$ and χ_{eff}), where each template is

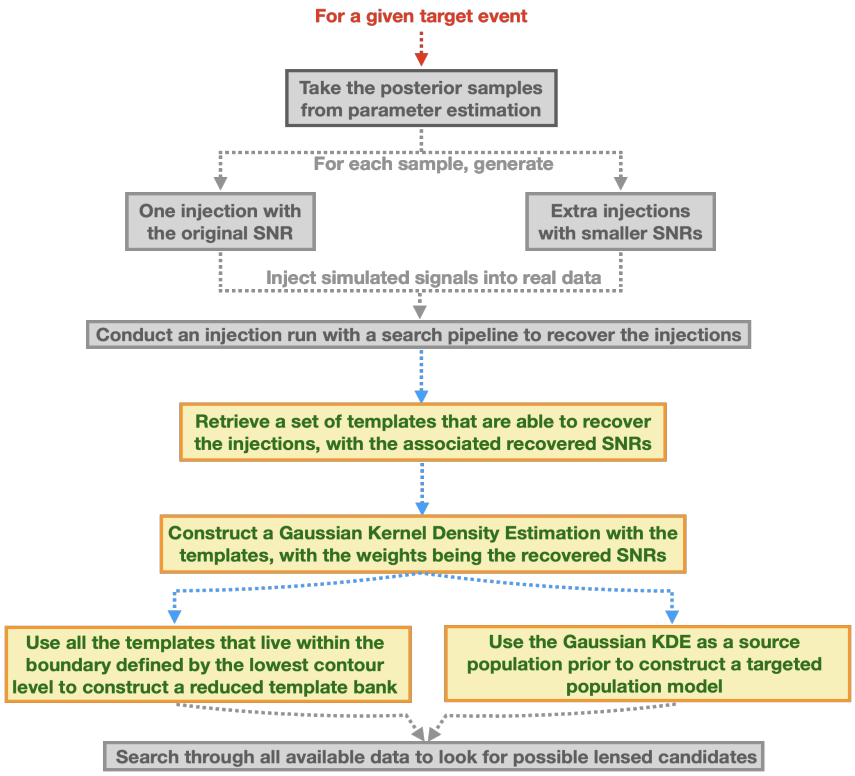

Figure 4.3: Workflow of TESLA-X. Grey elements represent steps directly inherited from the traditional TESLA workflow, while colored elements highlight new modules and modifications unique to TESLA-X

weighted by the SNR of the injections it recovers; weakly constrained spin measurements do not bias this step, because poorly measured spins lead to posterior samples that effectively reflect the prior, so the injections do not introduce artificial spin structure.

The KDE $f(\vec{\gamma})$ is evaluated at every template point $\vec{\theta}$ in the full bank and standard contour finding algorithms are used to determine characteristic KDE contour levels, as sketched in Fig. 4.1. The lowest contour is adopted as an empirical boundary in parameter space, and TESLA-X keeps all templates inside this boundary to construct the TESLA-X reduced bank. Because no templates are removed from within the selected region, the TESLA-X bank preserves the optimal spacing of the original bank in that subspace, ensuring that signals whose parameters lie inside the boundary are recovered with less than about three percent SNR loss and avoiding missed counterparts due to excessive mismatch.

Since the Gaussian KDE encodes the expected distribution of possible sub-threshold lensed counterparts, it is interpreted as an astrophysical source population prior. Following the likelihood ratio formalism described previously and in Reference [36], this KDE is used to construct the population term $P(\vec{\theta} \mid \text{signal})$ in the ranking statistic, with an explicit dependence on SNR so that templates further from the target posterior subspace are not over-penalized when they recover very weak signals. Figure 4.2 provides a schematic view of this targeted population model, and comparisons with a simple uniform mass model show that the targeted KDE based prior yields superior search sensitivity.

With both the TESLA-X template bank and the associated targeted population model in hand, `GstLAL` is run again to search the data for candidate sub-threshold lensed counterparts to the chosen target event. The

Properties	Injected super-threshold signal
UTC time	May 08 2022 11 : 06 : 00
GPS time	1336043178.397
Distance (Mpc)	2858.18
Primary mass $m_1^{\rm det}$	$70.08 M_\odot$
Secondary mass $m_2^{\rm det}$	$38.83 M_\odot$
Dimensionless spins	$\chi_{1x} = 0.182, \chi_{1y} = 0.182, \chi_{1z} = -0.0363,$ $\chi_{2x} = -0.113, \chi_{2y} = 0.132, \chi_{2z} = 0.116,$
Right ascension α	2.811
Declination δ	0.819
Inclination ι	2.513
Polarization Ψ	1.187
Waveform	IMRPhenomXPHMpseudoFourPN

Table 4.1: Summary of the properties of the injected super threshold GW signal MS220508a used in the simulation campaign, with all quantities given in the detector frame.

pipeline outputs a list of candidate triggers ranked by their likelihood ratios and associated false-alarm rates; these false-alarm rate values quantify the probability that a trigger is produced by noise and should be used only as a prioritization tool for follow up studies, not as a direct probability that a candidate is a genuine lensed counterpart. Future extensions will introduce an additional lensing specific likelihood term that incorporates explicit lensing information, enabling a more direct statistical assessment of whether a given candidate is truly a lensed image of the target.

4.3 Results

TESLA-X is built on the same core working principle as TESLA (See Section 4.1.1), differing only in how the reduced template bank is constructed and in the form of the assumed population model. For this reason, a new, dedicated simulation campaign is performed here with the specific goal of comparing the search performance and sensitivity of traditional TESLA against the proposed TESLA-X framework for identifying potential sub-threshold lensed GW signals.

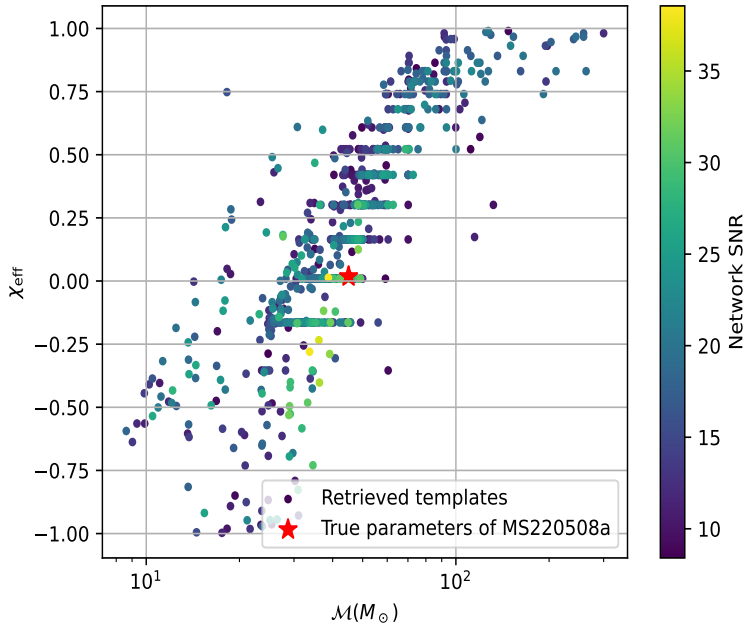

Figure 4.4: Templates activated by the simulated lensed injections of MS220508a in the injection campaign, shown in the chirp mass $\mathcal{M}_c$ versus effective spin $\chi_{\rm eff}$ plane. Marker colors indicate the corresponding network SNR of each injection

Injections	General	TESLA	TESLA-X
Total	7815	7815	7815
Found	6151	6169	6297
Found % change	-	+0.29%	+2.37%

Table 4.2: Number of injections recovered in the mock data search using the general template bank, the TESLA method, and the TESLA-X method, respectively.

4.3.1 Mock data generation

A 9.15 day long mock data set is first generated using Gaussian noise that has been recolored to match the publicly expected power spectral densities of O4 [202]. Continuous observation is assumed, with no downtime or vetoed segments applied to the data stream. Into this simulated data, a single super-threshold GW signal is injected, modeled with the IMRPhenomXPHMpseudoFourPN waveform [188], which incorporates precession and higher order modes. The intrinsic and extrinsic parameters of this injected signal are summarized in Table 4.1, and in what follows this reference injection is denoted MS220508a.

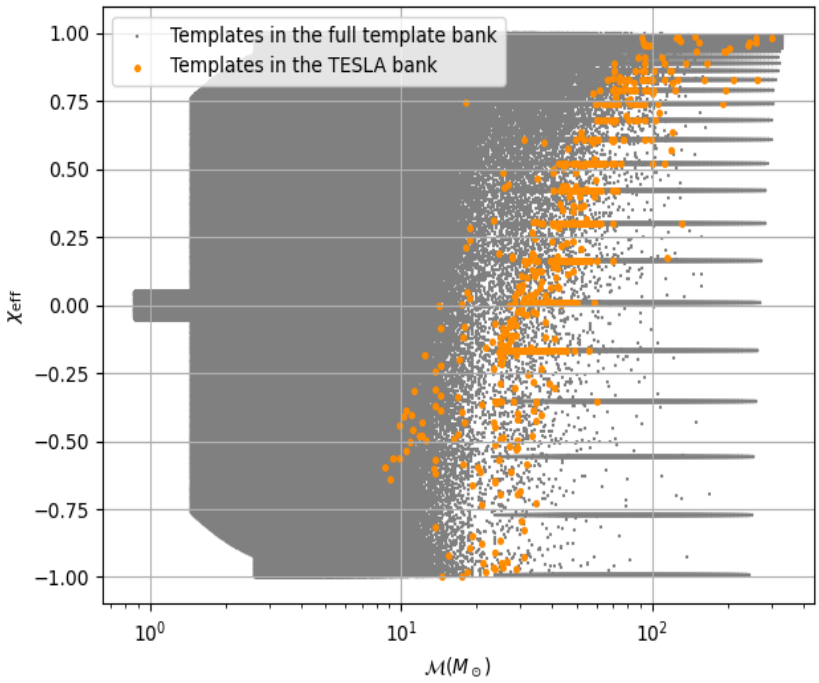

Figure 4.5: Templates from the full template bank shown in grey and from the TESLA reduced template bank shown in orange, displayed in the $\mathcal{M}_c$ versus $\chi_{\rm eff}$ space.

4.3.1.1 Performing a general search

A `GstLAL` search is then run on the mock data stream using the same configuration as in the O3 analysis reported in GWTC-3 [54]. Under this setup, the template bank is the same generic bank employed in Reference [18], covering the relevant compact binary parameter space. As expected, GstLAL recovers the injected signal MS220508a with the highest ranking statistic of all triggers, corresponding to a false-alarm rate of 2.972×10^{-35} Hz and an overall rank of one [54].

Next, parameter estimation for MS220508a is performed using the Bayesian inference library `bilby` [52], with sampling carried out under its default Markov chain Monte Carlo configuration. This analysis yields a posterior sample set for the source parameters, which then serves as the input for both the TESLA and TESLA-X search pipelines in the subsequent sub-threshold counterpart studies [52].

4.3.1.2 Performing an injection campaign

Posterior samples from the parameter estimation of MS220508a are used to generate 7815 simulated lensed injections according to Eq. 4.1, which are then added to the mock data and searched with `GstLAL` using the general template bank. The recovery results from this campaign provide the common input for both the TESLA and TESLA-X analyses.

In this setup, the general template bank successfully recovers 6151 of the 7815 injected signals, as listed in the "General" column of Table 4.2, with a total of 1008 distinct templates being rung up since a given template can recover multiple injections. The distribution of these templates in the $(\mathcal{M}_c, \chi_{\rm eff})$

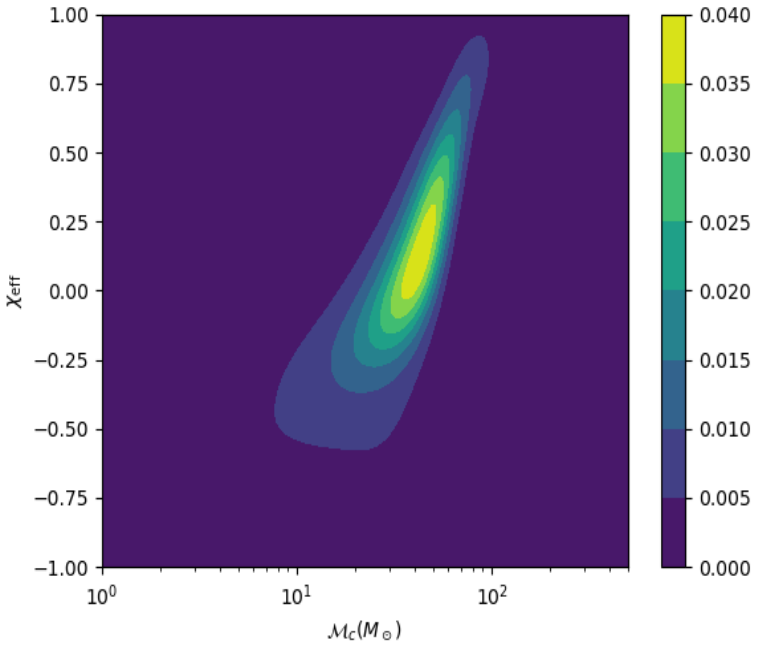

Figure 4.6: Contour map of the Gaussian kernel density estimation constructed for MS220508a in the $\mathcal{M}_c$ versus $\chi_{\rm eff}$ space, where the color scale denotes the probability density inferred from the KDE.

plane, colored by the network SNR of the associated recovered injections and shown in Fig. 4.4, aligns with expectations: templates near the true parameters of MS220508a are mostly linked to higher SNR, super-threshold injections, whereas templates farther away tend to recover lower SNR, sub-threshold injections.

4.3.1.3 Generating the TESLA and TESLA-X reduced template banks

All templates that are rung up in the injection campaign are retained to form the TESLA targeted template bank, with the full bank shown in grey and the TESLA-selected subset in orange in Fig. 4.5.

For TESLA-X, the same rung-up templates are used to build a Gaussian kernel density estimate KDE($\vec{\gamma}$), where each template is weighted by the network SNR of the injections it recovers and $\vec{\gamma}$ denotes the template labels, here the chirp mass $\mathcal{M}_c$ and effective spin $\chi_{\rm eff}$. The KDE is evaluated at every template in the full bank and visualized as a contour map, as shown in Fig. 4.6.

The lowest KDE contour is then adopted as the boundary of the targeted search region; for MS220508a this corresponds to a KDE value of 0.005. All templates inside this contour are kept to build the TESLA-X targeted template bank, illustrated in Fig. 4.7, where the full bank appears in grey, the TESLA-X templates in red, and the yellow curve marks the chosen KDE contour. The resulting TESLA-X bank for MS220508a contains 22136 templates.

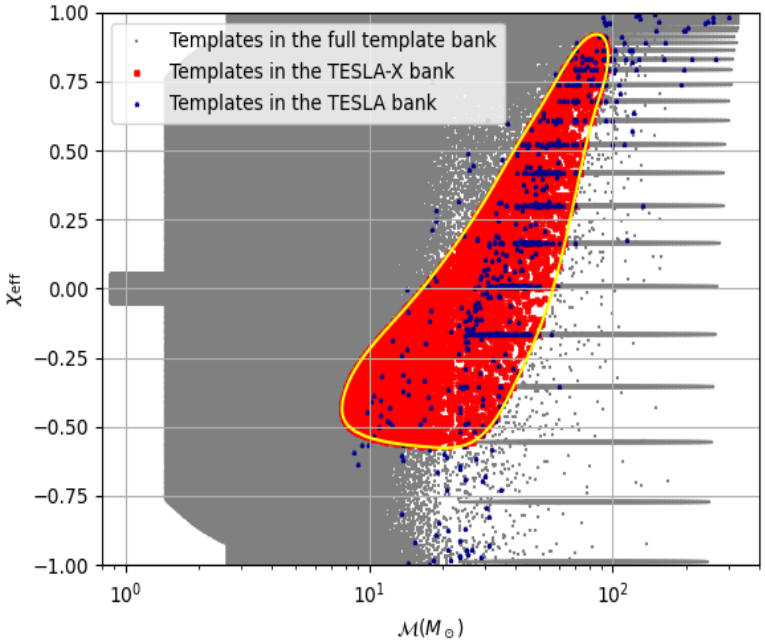

Figure 4.7: Templates from the full template bank shown in grey and from the TESLA-X reduced template bank shown in red, displayed in the $\mathcal{M}_c$ versus $\chi_{\rm eff}$ space. The TESLA-X template bank is obtained by retaining templates that lie within the lowest contour level of the Gaussian KDE shown in Figure 4.6, corresponding to a density value of 0.005, which is indicated by the yellow curve in this figure. For comparison, templates from the TESLA reduced template bank are also overlaid in blue.

4.3.1.4 Constructing the targeted population model for the TESLA-X reduced template bank

Using the KDE estimated for MS220508a and the TESLA-X reduced template bank, a targeted source population model is constructed following the procedure outlined in Section 4.2 and Reference [36]. For practical implementation, the Gaussian KDE is recomputed in the intrinsic parameter space of component masses and spins, $\vec{\gamma} = \{m_1, m_2, \chi_1, \chi_2\}$, rather than in terms of the derived quantities $\mathcal{M}_c$ and $\chi_{\rm eff}$, and this KDE then defines the population prior used in the TESLA-X likelihood ratio.

4.3.1.5 Performing a re-filtering to recover the lensed injections with the TESLA and TESLA-X methods

Two follow-up searches are then carried out on the same mock data set to recover the ensemble of lensed injections used in the injection campaign. One search employs the TESLA reduced template bank, while the other uses the TESLA-X reduced template bank together with the targeted population model, enabling a direct, like-for-like comparison of their performance. As summarized in Table 4.2, TESLA recovers slightly more injections than the original full template bank, corresponding to a gain of about 0.29%, whereas TESLA-X recovers an even larger number of lensed injections, improving on TESLA by 2.37%.

The sensitive range as a function of false-alarm rate for MS220508a-like signals is also compared for the full, TESLA, and TESLA-X template banks, as shown in Fig. 4.8. TESLA-X delivers the largest increase

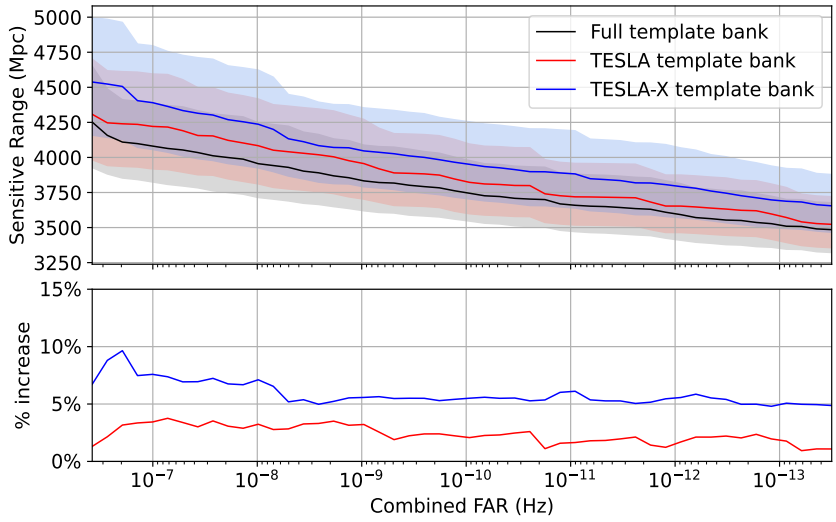

Figure 4.8: Top panel shows the sensitive range versus false-alarm rate for MS220508a-like signals obtained with the full template bank shown in black, the TESLA template bank shown in red, and the TESLA-X template bank shown in blue. The shaded region around each curve indicates the corresponding one sigma uncertainty. Bottom panel presents the fractional change in sensitive range as a function of false-alarm rate for the different template banks. Both the TESLA and TESLA-X banks yield improved sensitivity to MS220508a like lensed signals compared to the full bank. The TESLA-X curve consistently lies above the TESLA curve, indicating superior performance. The enhancement is most pronounced at false-alarm rates greater than or equal to 10^{-7}, Hz, corresponding to the sub-threshold regime, demonstrating the advantage of the TESLA-X method over the traditional TESLA approach.

in sensitivity to MS220508a like lensed signals, with the advantage being most pronounced at false-alarm rates $\gtrsim 10^{-7}$ Hz, i.e. in the sub-threshold regime. These results demonstrate that TESLA-X outperforms traditional TESLA in identifying candidate sub-threshold strongly-lensed GW signals.

4.4 Conclusions

The detectability of strongly-lensed GW signals can be diminished when lensing demagnifies the strain amplitude enough to push the signal below the usual detection threshold. Earlier work introduced the TESLA method [18], which recovers such strongly-lensed sub-threshold signals by building a targeted template bank informed by already detected events. The present study proposes an improved framework, TESLA-X, specifically designed to further enhance search sensitivity to strongly-lensed sub-threshold GWs.

TESLA-X starts by generating simulated lensed injections from posterior samples obtained in parameter estimation of a detected (super-threshold) event. These simulated counterparts are analyzed in an injection campaign with `GstLAL`, using a detection threshold corresponding to a false-alarm rate of 3.385×10^{-7} Hz to ensure statistically significant triggers. The injections that are successfully recovered are then used to construct both a targeted population model and a densely sampled reduced template bank, after which the data are reanalyzed with this targeted setup to search for candidate sub-threshold lensed signals.

The targeted population model is built in several stages. First, a Gaussian kernel density estimate is applied to the recovered injections in source parameter space $\vec{\gamma}$, producing a probability density function that describes the expected distribution of strongly-lensed sub-threshold counterparts [36]. In parallel, a targeted template bank is obtained by defining a boundary in parameter space, inferred from the recovered injections, and discarding all templates outside this boundary while retaining all templates inside. This yields a high density template bank concentrated on a restricted region of parameter space, and, unlike in TESLA, guarantees that any lensed counterpart within the boundary is recovered with less than about a three percent loss in SNR. The false-alarm rates and likelihood values $\mathcal{L}$ produced by the subsequent search are used to distinguish astrophysical signals from noise triggers, but they do not directly quantify the probability that a given candidate is a lensed counterpart of the target.

A dedicated simulation campaign compares three strategies: a search with the full template bank, the original TESLA method, and TESLA-X. Using a mock data set consisting of Gaussian noise plus an injected event labeled MS220508a, TESLA-X is found to recover more lensed injections associated with MS220508a than either the full bank search or TESLA. The analysis also reveals an improvement of up to roughly ten percent in the sensitive range when examining the sensitive range as a function of false-alarm rate, implying that TESLA-X can detect more distant or intrinsically weaker signals, consistent with the scaling in Eq. 4.1. These results show that TESLA-X delivers superior performance in searches targeting strongly-lensed GW signals.

Like TESLA, TESLA-X remains computationally expensive, since applying either framework to a single super-threshold event typically requires weeks of computation for the injection campaign. Future work will focus on accelerating TESLA-X, for example by optimizing the injection strategy or reducing the computational cost of repeated searches. Additional studies will also explore alternative parameterizations beyond chirp mass and effective spin for representing the targeted population, as outlined in the Appendix, with the aim of further improving both efficiency and sensitivity.

Chapter 5

Conclusion

The discovery of lensed GWs would be groundbreaking for both cosmology and GW astronomy, providing new probes of compact objects population, structure formation, and fundamental physics. These prospects have strongly motivated the development of dedicated GW lensing analyses. Nevertheless, the detection of lensed GW signals remains technically and statistically challenging. Addressing this problem across multiple stages, from detection to inference, this thesis outlines a coherent pathway toward the robust identification and characterization of lensed GWs in both the wave-optics and strong-lensing regimes.

After an introduction in Chapter 1, we systematically examine how wave-optics lensing affects the detectability of lensed GWs in matched-filtering searches by carrying out the first dedicated simulation campaign using the matched-filtering pipeline `GstLAL`.Contrary to the conventional expectation that larger signal amplitude lead to bettwe signal detectability, we find that highly magnified lensed GWs, even with amplitudes an order of magnitude larger, can experience a substantial reduction in detectability. This occurs because wave-optics distortions lower the match with templates that do not include lensing effects. While our results indicate that the current GW catalog constructed from matched-filtering searches is incomplete with respect to lensed signals, they also lead to several insights for the GW lensing community. First, incorporating lensing effects into template bank construction is necessary to recover the most distorted GW events. Second, existing constraints on compact lenses and estimates of wave-optics lensing rates need to be revisited. Third, our findings motivate systematic analyses of GW candidates identified by coherent wave burst pipelines, whose model independent nature makes them naturally sensitive to distorted signals.

In Chapter 3, we develop `DINGO-lensing` within a simulation-based inference framework. This approach enables rapid and scalable parameter inference without relying on computationally expensive traditional Monte Carlo sampling, while retaining comparable accuracy and statistical consistency. By performing a detailed analysis of GW231123, an event reported to show the strongest support for GW lensing to date, we provide the first quantitative assessment of the statistical significance of lensing based on waveform distortions. This study demonstrates that simulation-based inference can be used to discover lensed GWs directly from their distorted signals, marking an important step toward robust lensing discoveries in GW astronomy.

In Chapter 4, we introduce the TESLA-X method for searching strongly-lensed GW signals that lie below

the standard detection threshold. By exploiting information from a corresponding super-threshold event, we construct an reduced targeted template bank and an associated targeted population model through a dedicated injection campaign using a matched-filtering search pipeline to look for their faint lensed counterparts. This approach extends and improves upon the traditional TESLA sub-threshold search by increasing the effective detection horizon and sensitivity to lensed signals. TESLA-X has been continuing to contribute to searches for GW lensing signatures in GWTC-4.0 and represents an important step toward strengthening statistical confidence in the first detection of strongly-lensed GWs.

This thesis investigates the discovery of GW lensing as a problem that spans both matched-filtering searches and statistical inference. The results presented here establish GW lensing data analysis as a frontier, in which lensing effects can challenge standard search pipelines and inference methodologies. The strategies developed in this thesis provide a foundation for future discoveries of GW lensing, highlighting that advances in detector sensitivity must be accompanied by corresponding progress in data analysis techniques.

Chapter 6

Appendices

6.1 Training and validation of O5 nonlensed network

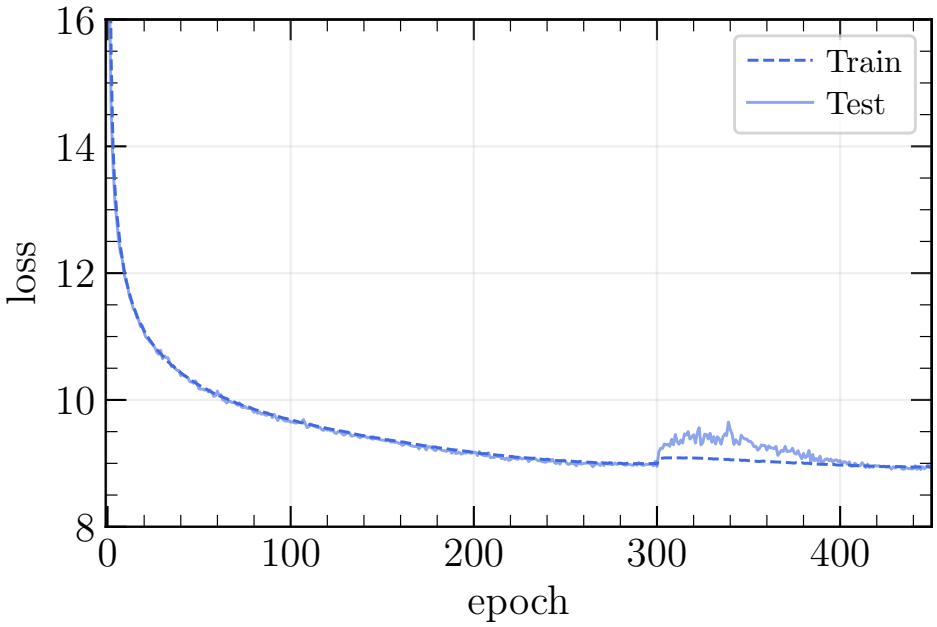

Figure 6.1: The loss as a function of the number of training epochs of the nonlensed network, shown for both the training set (dashed) and testing set (solid).

The nonlensed network is trained using the same setup and prior distributions as in the main analysis and Table 3.1, except that all lensing parameters are removed from the prior, and the architecture again contains about 10^8 trainable and 10^7 fixed parameters. Figure 6.1 displays the evolution of the loss during training, where the training loss (dashed) and validation loss (solid) remain closely aligned, indicating good generalization and no signs of overfitting. Both curves decrease smoothly and monotonically, demonstrating stable and well controlled optimization, with a clear change in slope near epoch 300 when Stage 1 begins and previously frozen layers are unfrozen.

After roughly 400 epochs, the loss flattens, signaling convergence and diminishing returns from additional

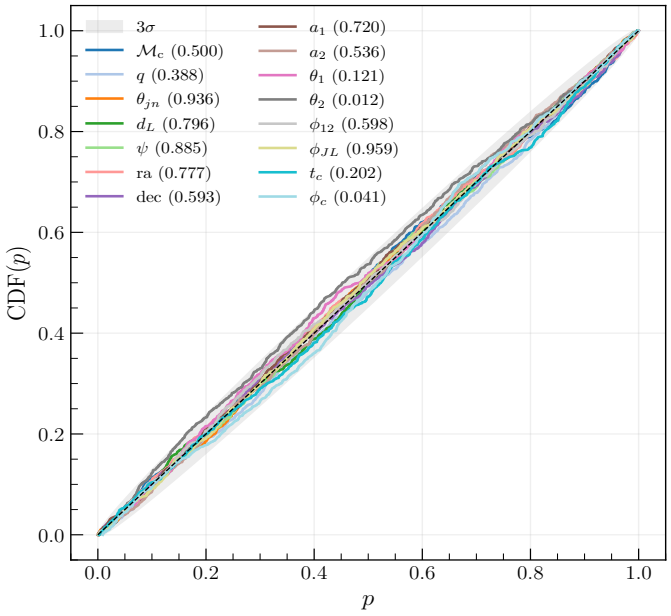

Figure 6.2: Probability–probability plot for the nonlensed neural posterior estimation model constructed using 10^3 nonlensed injections and no importance sampling. For each injection, a posterior distribution is generated and the percentile of the injected parameter within that posterior is evaluated. Each colored curve represents the cumulative distribution function of an individual parameter, with the corresponding Kolmogorov-Smirnov test p-values reported in the legend.

training, so the run is stopped at epoch 450 once no further improvement of the validation loss is observed. The full training procedure for the nonlensed model takes approximately 10 days on a single NVIDIA A100 PCIe GPU, consistent with reported `DINGO` training times for similar architectures.

Figures 6.2 shows the probability–probability plot constructed from the posterior distributions of the nonlensed injection set, where the black diagonal denotes the ideal case in which $x\%$ of true parameter values fall within the corresponding x percent credible intervals. The colored curves for individual parameters closely follow this diagonal, showing that the inferred posteriors are statistically well calibrated and that the inference network accurately represents the underlying parameter space. Kolmogorov–Smirnov test p values, reported in the legend, further support the robustness of the sampling procedure for the nonlensed model.

Figure 6.3 shows the distribution of the sampling efficiency ϵ obtained from the nonlensed inference network. We carry out inference on 10^3 nonlensed injections, drawing 10^5 posterior samples from the nonlensed network for each injection. The top panel displays the histogram of $\log_{10}(\epsilon)$, revealing that the majority of samples have efficiencies in the range $10^{-2.5}$ to 10^{-1}, with a low efficiency tail extending to smaller values. The bottom panel maps the efficiency across the source parameter space, where each point represents an injection and is colored by $\log_{10}(\epsilon)$ in the $(\mathcal{M}_c, d_L)$ plane. A weak correlation is observed, with higher efficiencies tending to occur for systems with larger chirp masses and greater luminosity distances. This indicates that the nonlensed network achieves more effective posterior reconstruction for lower signal to noise ratio and more massive binaries. Overall, the network demonstrates robust and consistent performance across the population, with variations in ϵ largely driven by differences in signal

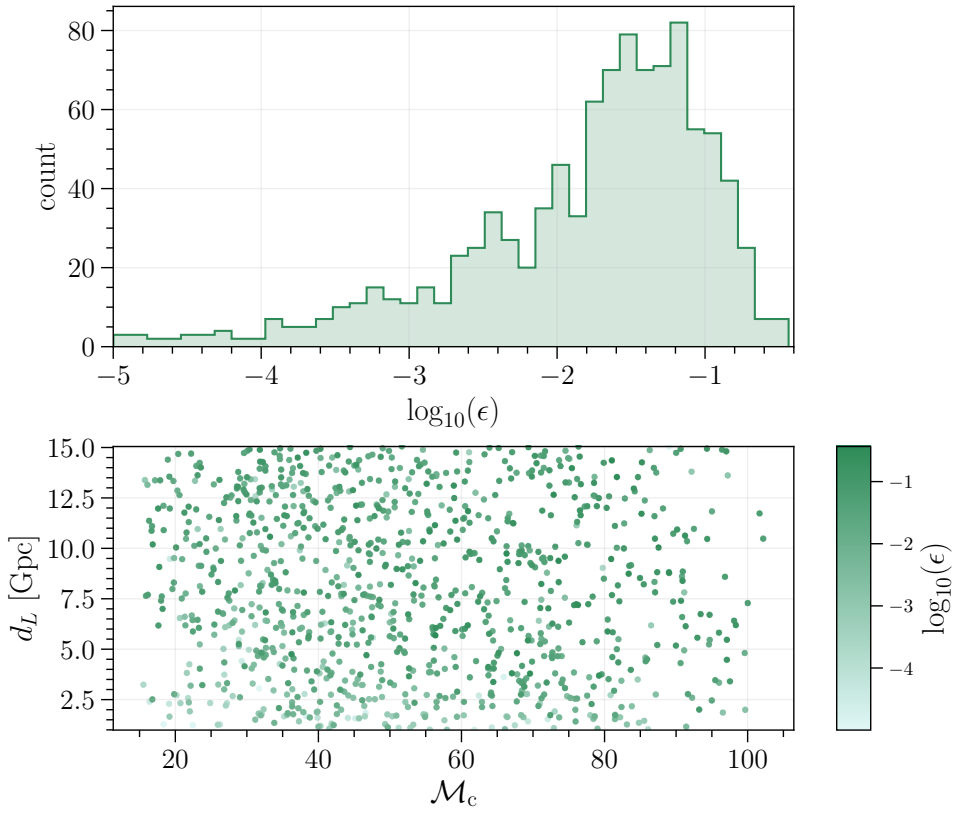

Figure 6.3: Distribution of the sampling efficiency ϵ of 10^3 nonlensed injections for the nonlensed network. 10^5 samples are drawn from the nonlensed network. The top panel shows the histogram of $\log_{10}(\epsilon)$, illustrating that most posterior samples have efficiencies between $10^{-2.5}$ and 10^{-1}. The bottom panel presents the relation between luminosity distance d_L and chirp mass $\mathcal{M}_c$, color-coded by $\log_{10}(\epsilon)$. The results indicate that lower-efficiency samples are more prevalent at small d_L and moderate $\mathcal{M}_c$, while the overall spread remains broad across the parameter space.

strength rather than systematic inference errors.

Figure 1.5 displays a corner plot comparing posterior distributions from the `DINGO` nonlensed model (solid green contours) to those from the reference `bilby` analysis (dashed gray contours) for the same lensed injection considered in Sec. 3.3.1.3, using injected values and priors listed in Table 3.1. The one dimensional and two dimensional marginalized posteriors agree closely between `DINGO` and `bilby` for both intrinsic and extrinsic parameters, demonstrating that the neural posterior estimator reproduces full Bayesian inference results, with a sampling efficiency of about 0.3 percent in this example

6.2 Bayes factor dependence on the sampling efficiency

Sampling efficiency after importance sampling is used to quantify how well the lensed and nonlensed networks reconstruct the target posteriors across the validation sets.

For the lensed network, inference is carried out on 10^3 lensed injections, drawing 10^5 samples per event from the trained model, with the resulting sampling efficiencies summarized in Fig. 3.3. The top panel shows a histogram of $\log_{10}(\epsilon)$ spanning roughly 10^{-1} to 10^{-4}, while the middle panel displays $\log_{10}(\epsilon)$ in the $(\mathcal{M}_c, d_L)$ plane, revealing higher efficiencies at larger luminosity distance, consistent with broader, lower–SNR posteriors being easier to match to the target distribution. The bottom panel maps $\log_{10}(\epsilon)$ over the lens parameter space and shows no clear dependence of sampling efficiency on the lensing parameters, indicating that performance is largely insensitive to lens configuration.

Figure 6.3 presents the corresponding analysis for the nonlensed network, again using 10^3 injections and 10^5 samples per event. The upper panel shows a histogram of $\log_{10}(\epsilon)$ concentrated between about $10^{-2.5}$ and 10^{-1} with a tail toward lower efficiencies, while the lower panel shows $\log_{10}(\epsilon)$ across the $(\mathcal{M}_c, d_L)$ plane, where a mild trend toward higher efficiency at larger chirp mass and distance suggests more effective reconstruction for lower–SNR and more massive systems. Overall, the nonlensed network exhibits stable performance across the population, with variations in ϵ dominated by differences in signal strength rather than systematic biases in the inference.

The lensing Bayes factor $\log_{10} \mathcal{B}_{\text{lens}}$ serves as the primary metric for evaluating lensed candidates, and its performance is validated by comparing `DINGO-lensing` against `bilby` on a set of randomly selected lensed and nonlensed simulations spanning a range of Bayes factors and sampling efficiencies, restricted to cases with $\log_{10} \mathcal{B}_{\text{lens}} > 0$. For each event analyzed with `DINGO-lensing`, 5×10^4 posterior samples are drawn, and the resulting Bayes factors, summarized in Table 6.1, show agreement with `bilby` within statistical uncertainties whenever the sampling efficiency satisfies $\epsilon > 10^{-4}$. Simulations with lower efficiencies are therefore discarded, leaving a final set of 2,561 lensed and 10,858 nonlensed events used in subsequent analyses.

During validation, some simulated events with mass ratios near the lower edge of the training prior, $q = 0.125$, exhibit unreliable evidence estimates from the networks, consistent with the known difficulty of learning sharply peaked posteriors close to prior boundaries in neural posterior estimation. To address this, both the foreground and background validation samples are conservatively restricted to systems with $q > 0.3$, a regime in which the agreement between `DINGO-lensing` and `bilby` remains consistently good, as shown in the preceding comparisons. The detailed parameter choices for the resulting validation subset are summarized in Table 6.2.

6.3 Distance away from the nonlensing hypothesis

For the adopted two-image interference lens model (See Section 1.4.1.6), a nonlensed signal corresponds to the point $\Delta t = \mu_{\text{rel}} = 0$ in lens parameter space. The degree of inconsistency with this hypothesis can therefore be quantified by how far the lensing posterior samples lie from this reference point, a quantity

denoted by Δ_{lens}.

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

Table 6.1: Comparison of the lensing Bayes factor $\log_{10} \mathcal{B}_{\text{lens}}$ computed with `DINGO-lensing` and `bilby` for nonlensed and lensed simulated event with different sampling efficiencies ϵ. We report both the sampling efficiency obtained with the lensed (ϵ_{lens}) and nonlensed ($\epsilon_{\text{nonlens}}$) networks.

For this distance to be interpretable, it must be rescaled by the intrinsic uncertainties of the inferred parameters. Approximating the posterior as Gaussian, the distance measured in units of the standard deviation is defined following Ref. [196].

$$D(\theta_1, \theta_2) = \sqrt{(\theta_1 - \theta_2))^T (C_1 + C_2)^{-1} (\theta_1 - \theta_2)}\,, \tag{6.1}$$

where θ_1 and θ_2 denote the parameters over which the distance is evaluated, and C_1 and C_2 are the corresponding covariance matrices. In this case, $\theta_1 = \Delta t, \mu_{\rm rel}$ and $\theta_2 = 0, 0$, which defines the reference point corresponding to the nonlensed hypothesis.

$$\mathcal{D}_{\rm nonlens} = D(\{\Delta t, \mu_{\rm rel}\}, 0)\,. \tag{6.2}$$

Because the prior on $\{\Delta t, \mu_{\rm rel}\}$ is bounded from below at zero, the corresponding lensing distance satisfies $\Delta_{\rm lens} \gtrsim 1$ for nonlensed events consistent with the prior support. For posteriors that deviate from Gaussianity, the Gaussian approximation tends to underestimate the true separation in parameter space, yielding smaller inferred distances, so $\Delta_{\rm lens}$ acts as a conservative statistic for flagging lensed candidates.

A key advantage of $\Delta_{\rm lens}$ relative to the lensing Bayes factor $\mathcal{B}_{\rm lens}$ is that it does not depend on importance sampling and can be computed directly from the neural posterior returned by `DINGO-lensing`, providing a fast and robust diagnostic of lensing signatures.

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

Table 6.2: Parameters for the simulated injections in Table 6.1.

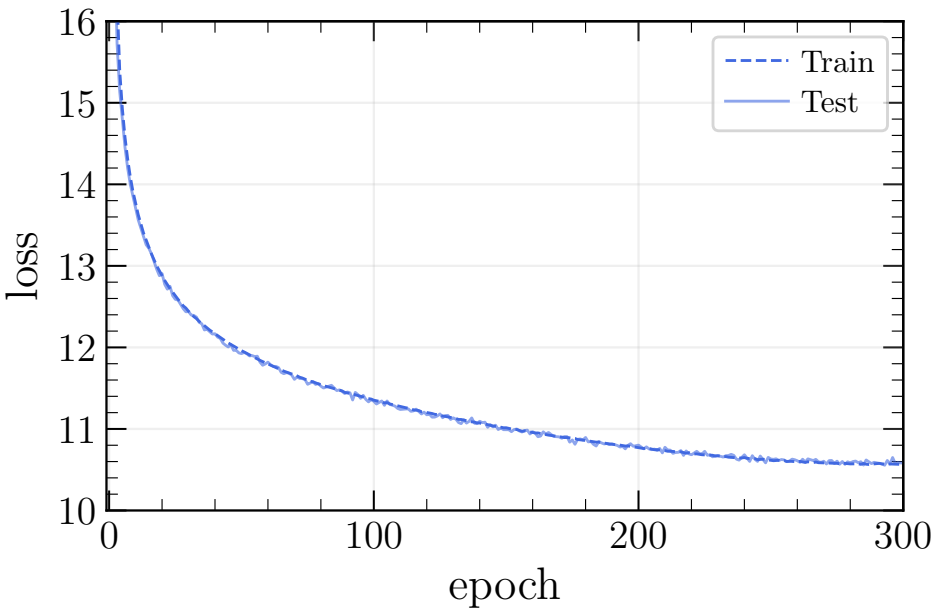

Figure 6.4: The loss as a function of the number of training epochs for the lensed network targeted for GW231123, shown for the training set as dashed curves and for the testing set as solid curves.

6.4 Training and validation of lensed-GW231123 network

Figure 6.4 illustrates the evolution of the loss as a function of training epoch for the targeted network for GW231123. The training loss shown with a dashed curve and the test loss shown with a solid curve track each other closely throughout the training process, indicating the absence of overfitting. Both curves evolve smoothly, demonstrating stable and well behaved optimization. After approximately 200 epochs, the loss reaches a plateau, suggesting that the network has saturated the information content of the training data.

Figure 6.5 presents the probability–probability plot built directly from the posteriors produced by the lensed network, without any importance sampling. The black diagonal indicates ideal calibration, where a given fraction of true parameter values falls within the corresponding credible intervals, and the colored curves for individual parameters closely track this line, showing that the network produces well calibrated posterior distributions. This behavior confirms that the sampling procedure operates as intended, a conclusion further supported by the Kolmogorov–Smirnov test p values listed in the legend, and the small deviations seen for the lensing parameters Δt and $\mu_{\rm rel}$ remain within the expected 3σ statistical fluctuations, indicating robust performance even in the presence of lensing.

Figure 6.6 shows the distribution of sampling efficiencies ϵ obtained after importance sampling span a wide range, from about 10^{-4} up to 10^{-1}, and show no clear correlation with either the binary source parameters or the lensing parameters across the validation set.

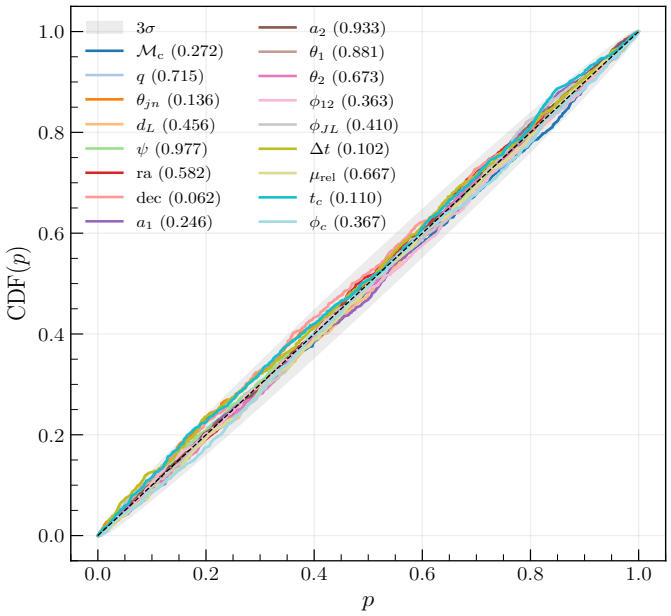

Figure 6.5: Probability–probability plot for the lensed neural posterior estimation model targeted for GW231123, constructed from 10^3 lensed injections without applying importance sampling. For each injection, a posterior is drawn from the model and the percentile rank of the injected parameter is computed. Each colored curve represents the cumulative distribution function of a given parameter, as indicated in the legend. The corresponding Kolmogorov–Smirnov test p values are also listed in the legend.

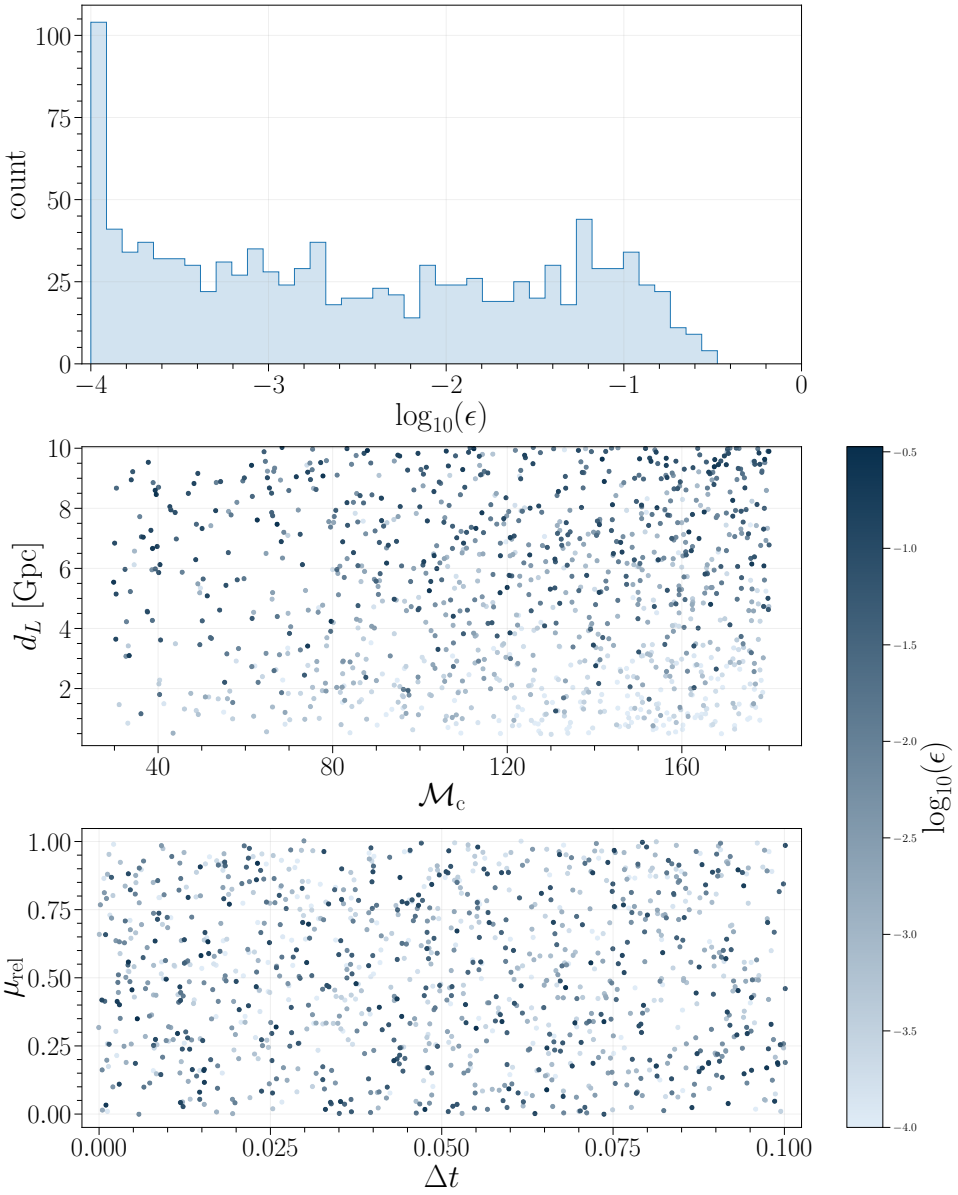

Figure 6.6: Distribution of the sampling efficiencies ϵ of 10^3 lensed injections for the lensed network targeted for GW231123. 10^5 samples are drawn for each injections The top panel shows the histogram of $\log_{10}(\epsilon)$, with most samples distributed between 10^{-4} and 10^{-1}. The middle panel illustrates the dependence of ϵ on the source parameters, where each point represents an injection colored by $\log_{10}(\epsilon)$ in the $(\mathcal{M}_c, d_L)$ plane. A weak trend is observed in which higher $\mathcal{M}_c$ and larger d_L correspond to an improved sampling efficiency. The bottom panel shows the variation of ϵ with the lensing parameters $(\Delta t, \mu_{rel})$, indicating no strong correlation but a broad spread consistent with the added complexity of the lensing parameter space.

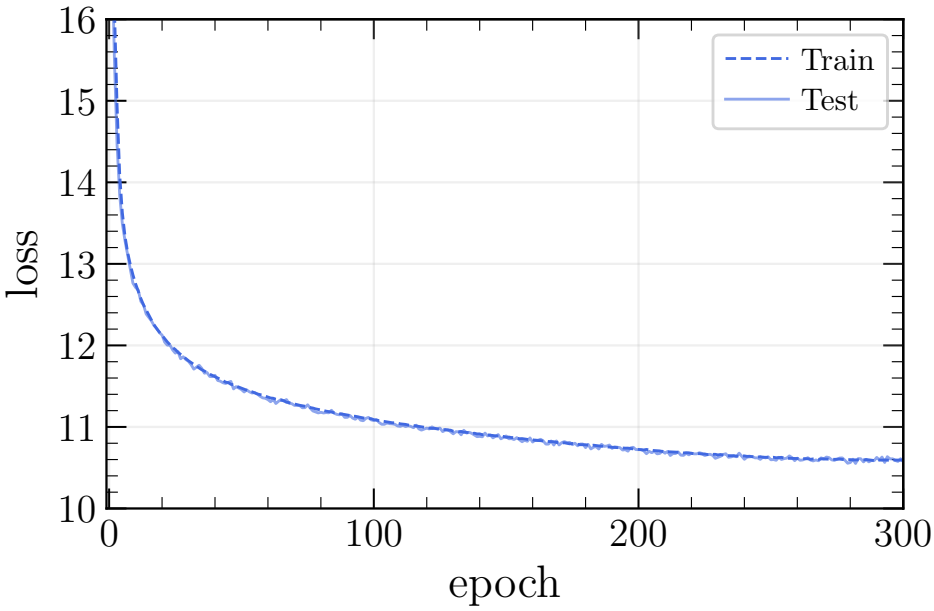

Figure 6.7: The loss as a function of the number of training epochs for the nonlensed network targeted for GW231123, shown for the training set as dashed curves and for the testing set as solid curves.

6.5 Training and validation of nonlensed-GW231123 network

The nonlensed network for GW231123 is trained with the same configuration and prior choices as used in the main analysis and listed in the fourth column of Table 3.1, with the exception that all lensing parameters are excluded from the prior. The network architecture again comprises approximately 10^8 trainable parameters and 10^7 fixed parameters. Figure 6.7 shows the training history, where the dashed training loss and solid validation loss track each other closely, indicating good generalization without evidence of overfitting. Both losses decrease smoothly and monotonically, reflecting stable and well controlled optimization, with a distinct change in slope around epoch 300.

Figure 6.8 shows the probability-probability plot constructed directly from the posteriors generated by the lensed network, without applying importance sampling. The black diagonal denotes ideal calibration, where a given fraction of true parameter values lies within the corresponding credible intervals, and the colored curves for individual parameters closely follow this reference, demonstrating that the network yields well calibrated posterior distributions. This behavior confirms that the sampling procedure functions as intended, a conclusion further supported by the Kolmogorov-Smirnov test p values reported in the legend.

Figure 6.9 presents the distribution of the sampling efficiency ϵ obtained from the nonlensed inference network. We perform inference on 10^3 nonlensed injections and draw 10^5 posterior samples from the nonlensed network for each one. The top panel shows the histogram of $\log_{10}(\epsilon)$, indicating that most samples exhibit efficiencies between $10^{-2.5}$ to 10^{-1}, with a small tail extending toward lower values. The bottom panel illustrates how the efficiency varies across the source parameter space: each point corresponds to an injection and is color-coded by $\log_{10}(\epsilon)$ in the $(\mathcal{M}_c, d_L)$ plane. A mild trend is

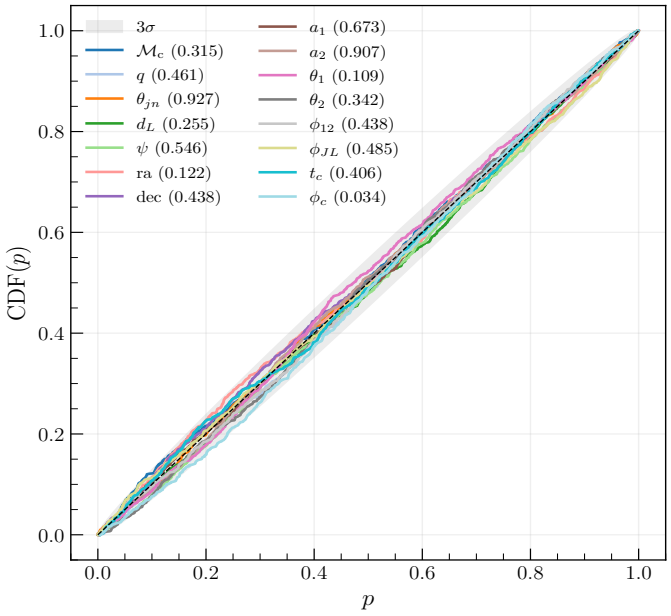

Figure 6.8: Probability–probability plot for the nonlensed neural posterior estimation model targeted for GW231123, constructed from 10^3 lensed injections without applying importance sampling. For each injection, a posterior is drawn from the model and the percentile rank of the injected parameter is computed. Each colored curve represents the cumulative distribution function of a given parameter, as indicated in the legend. The corresponding Kolmogorov–Smirnov test p values are also listed in the legend.

observed, where higher efficiencies are typically associated with systems of larger chirp mass and greater luminosity distance. This suggests that the nonlensed network reconstructs posteriors more effectively for lower–signal-to-noise ratio and more massive binaries. Overall, the network maintains stable and reliable performance across the entire population, with fluctuations in ϵ mainly reflecting differences in signal strength rather than systematic biases in inference.

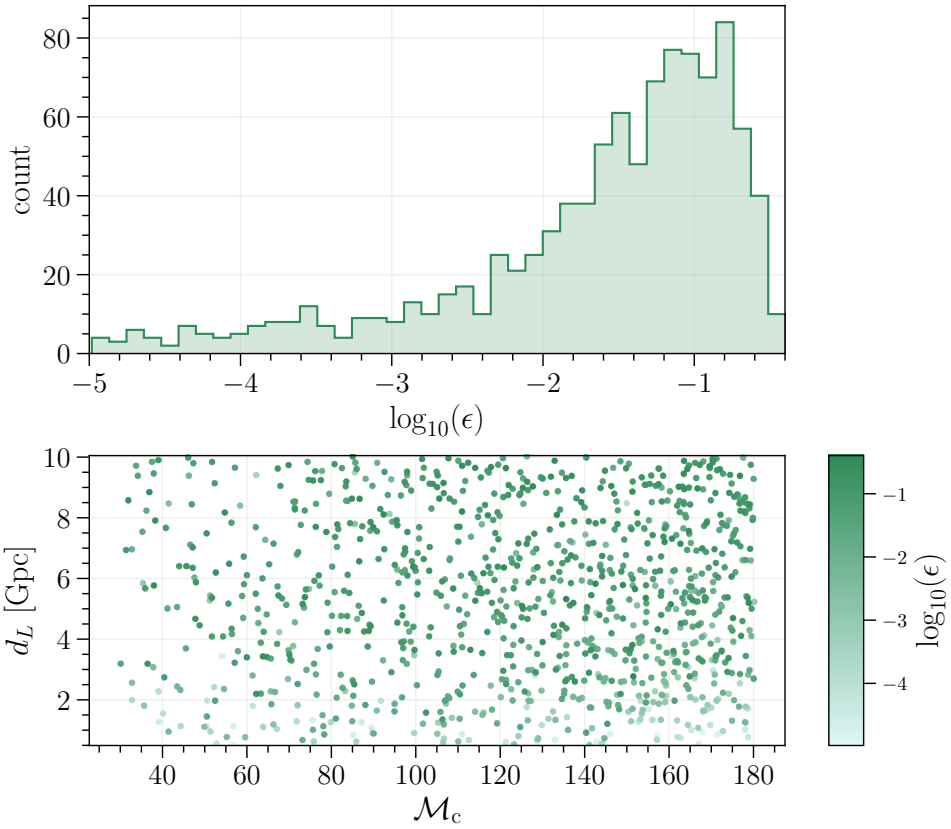

Figure 6.9: Distribution of the sampling efficiency ϵ of 10^3 nonlensed injections for the nonlensed network targeted for GW231123. 10^5 samples are drawn from the nonlensed network. The top panel shows the histogram of $\log_{10}(\epsilon)$, illustrating that most posterior samples have efficiencies between $10^{-2.5}$ and 10^{-1}. The bottom panel presents the relation between luminosity distance d_L and chirp mass $\mathcal{M}_c$, color-coded by $\log_{10}(\epsilon)$. The results indicate that lower-efficiency samples are more prevalent at small d_L and moderate $\mathcal{M}_c$, while the overall spread remains broad across the parameter space.

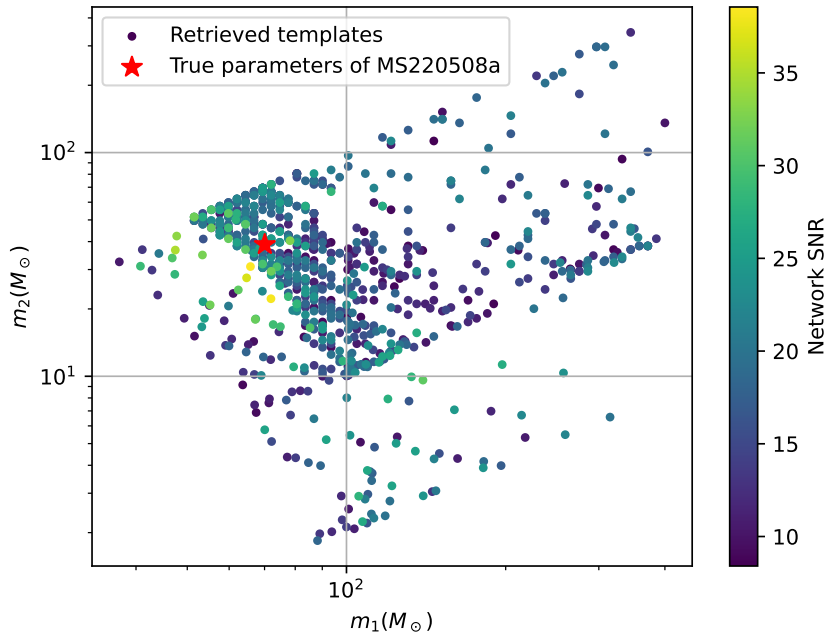

Figure 6.10: Rung-up templates by the simulated lensed injections of MS220508a in the injection campaign, plotted on the component masses m_1-m_2 parameter space. The colors of the markers represent the network SNRs of the associated injections.

6.6 Choosing a suitable pair of parameters to construct the Gaussian KDE

In the main analysis, the Gaussian kernel density estimate is constructed in the two dimensional space defined by chirp mass $\mathcal{M}_c$ and effective spin $\chi_{\rm eff}$, which are convenient combinations of the component masses and spins that are relatively well constrained by current GW data. It has been suggested, however, that working directly in the component mass plane spanned by (m_1, m_2) may yield a more natural or informative representation for some applications. This appendix presents preliminary investigations exploring the use of the (m_1, m_2) parameterization as an alternative basis for building the KDE and the associated targeted population model.

6.6.1 Redoing the TESLA-X analysis for MS220508a with the Gaussian KDE constructed in the m_1-m_2 space

In this subsection, the TESLA-X analysis is repeated with the Gaussian kernel density estimation constructed directly in the component mass parameter space (m_1, m_2), instead of in the $(\mathcal{M}c, \chi$eff$)$ space used in the main text.

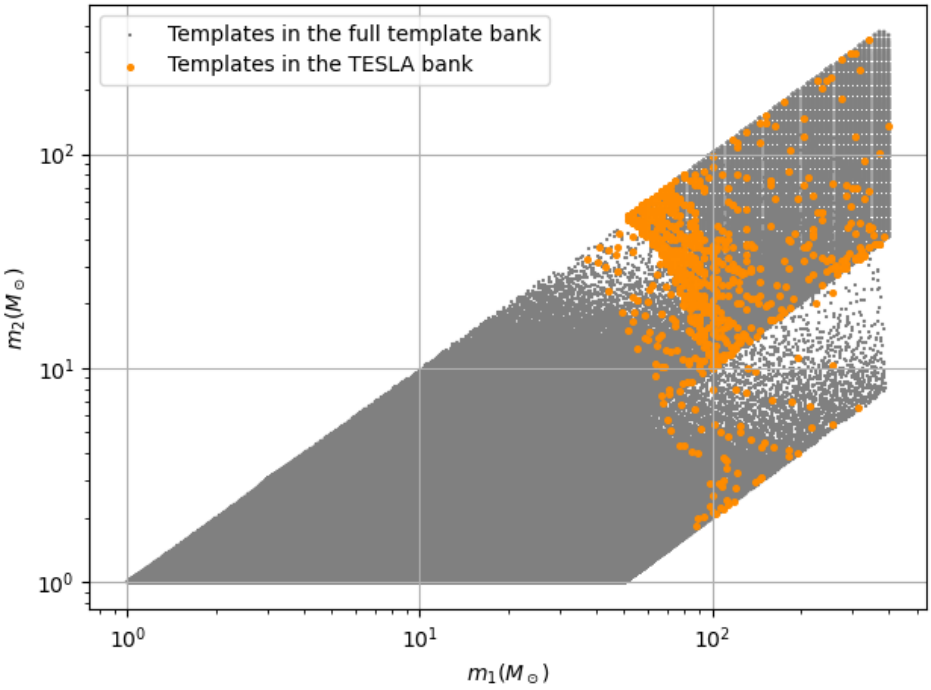

Figure 6.11: The templates in the full template bank (in grey) and in the TESLA reduced template bank (in orange), plotted in the component masses m_1-m_2 space.

6.6.1.1 Constructing the TESLA-X bank and population model

The analysis begins from the same injection run results used for MS220508a in the main text. Figure 6.10 shows the rung up templates projected onto the component mass plane (m_1, m_2), with marker colors indicating the network SNRs of the associated injections, thereby illustrating how higher SNR injections cluster nearer the true source masses.

Although the parameterization is changed, the TESLA targeted template bank itself remains identical to that in the main text, because it is still defined as the set of all templates that were rung up by the injections during the campaign. For completeness, the same TESLA bank is therefore re-plotted in the (m_1, m_2) space in Fig. 6.11, with the full bank shown in grey and the TESLA templates highlighted in orange.

For TESLA-X, the rung up templates from the injection campaign are used to construct a Gaussian kernel density estimate KDE($\vec{\gamma}$) in the component mass space, with $\vec{\gamma} = \{m_1, m_2\}$ and each template weighted by the network SNR of the injections it recovers. The KDE value is then evaluated at every template in the full bank and displayed as a contour map, as illustrated in Fig. 6.12. The lowest contour level, corresponding to a KDE value of 2×10^{-5}, is adopted as the boundary defining the targeted region; all templates within this boundary are retained to form the "TESLA-X (component mass) bank". Figure 6.13 shows the full bank in grey and this TESLA-X (component mass) subset in red, with the yellow curve marking the chosen contour, and the resulting bank contains 159043 templates.

As in the main analysis, the KDE estimated for MS220508a in the (m_1, m_2) space, together with the TESLA-X (component mass) bank, is then used to construct a targeted population model following the

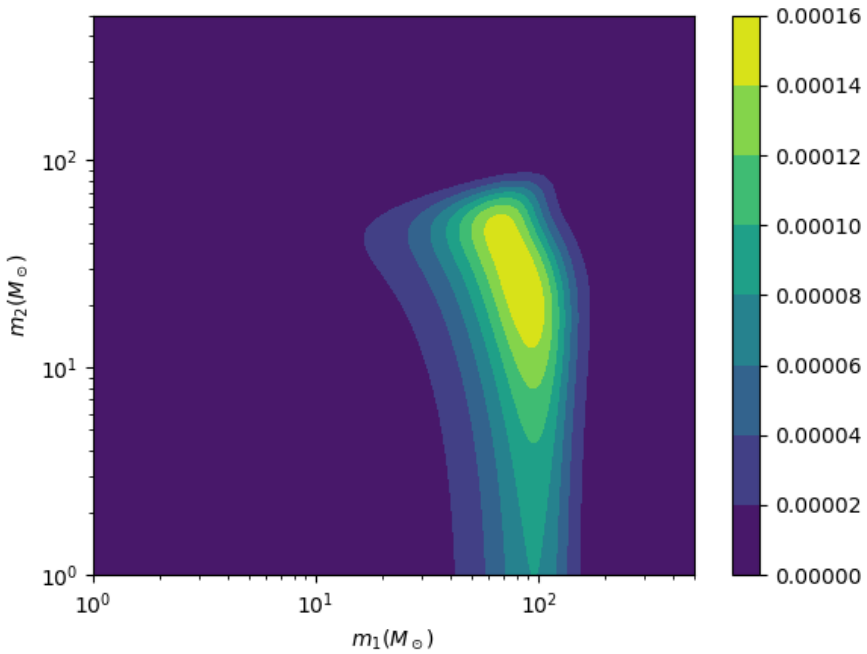

Figure 6.12: The contour map of the Gaussian Kernel Density Estimation function obtained for MS220508a in the m_1 - m_2 space. The colors represent the probability density estimated by the KDE function.

same procedure described in Section 4.2 and Reference [36], providing an alternative, component mass based representation of the expected sub-threshold lensed counterpart population.

6.6.1.2 Trying to recover the lensed injections with the "TESLA-X (main) bank" and "TESLA-X (component mass) bank"

To assess the relative performance of the TESLA-X main bank and the TESLA-X component mass bank for MS220508a, an additional search is conducted using the TESLA-X component mass bank together with its targeted population model, applied to the same mock data set and aiming to recover the same ensemble of lensed injections used in the injection campaign. The corresponding recovery statistics are reported in Table 6.3.

This configuration is found to perform worse than the search with the full template bank, showing a decrease of about 0.3% in the number of recovered lensed injections. The result is consistent with expectations, since the TESLA-X component mass bank includes almost six times as many templates as the TESLA-X main bank, which substantially increases the trials factor, enlarges the effective noise background, and thereby degrades the ability of the search to recover the simulated lensed injections.

Examining the sensitive range as a function of false-alarm rate for MS220508a like signals with three configurations, the full template bank, the TESLA-X main bank, and the TESLA-X component mass bank, shows that the TESLA-X component mass bank underperforms even the full bank, as illustrated in Fig. 6.14. For this case, the TESLA-X main bank provides the largest sensitive range at a given false-alarm

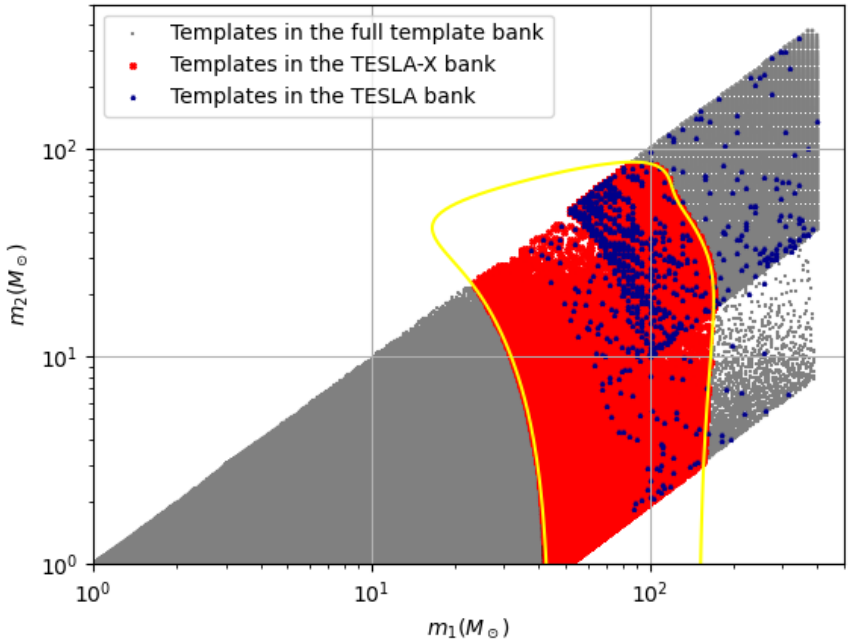

Figure 6.13: The templates in the full template bank (in grey) and in the "TESLA-X (component mass) bank" (in red), plotted in the m_1-m_2 space. The "TESLA-X (component mass) bank" is constructed by keeping templates that fall within the lowest contour level from the Gaussian KDE function shown in Figure 6.12, i.e. 0.00002, plotted as a yellow curve in this figure. For easy comparison, we also plot the templates in the TESLA reduced template bank in the same figure (in blue).

rate, while the component mass based bank yields a smaller range over most of the false-alarm rate domain. This comparison indicates that, at least for MS220508a, constructing the Gaussian kernel density estimation in terms of chirp mass $\mathcal{M}_c$ and effective spin $\chi_{\rm eff}$ offers a more effective parameterization for the TESLA-X framework than using the (m_1, m_2) component mass space.

6.6.2 Analyzing another mock event MS220510ae with the TESLA-X method

MS220508a corresponds to a comparatively low chirp mass system of approximately 45.0 $M_\odot$. To assess the robustness of the TESLA-X construction and the choice of parameters used in the Gaussian KDE, it is informative to repeat the analysis for a mock event occupying a different region of parameter space. In this subsection, we therefore replicate the procedure of the previous subsection using a second super-threshold mock event, MS220510ae. The mock data configuration is identical to that adopted in the main text, and the injected source parameters are summarized in Table 6.4.

Following the same workflow, we first perform Bayesian parameter estimation on MS220510ae and use the resulting posterior samples to generate a set of lensed injections. These injections are then searched using the full template bank, producing the injection recovery results that are subsequently used to construct two reduced banks, denoted as the TESLA-X main bank and the TESLA-X component mass bank, together with their associated targeted population models. Out of a total of 8099 lensed injections, 5047 are

Injections	General	TESLA-X (main)	TESLA-X (component mass)
Total	7815	7815	7815
Found	6151	6297	6132
Found % change	-	+2.37%	−0.3%

Table 6.3: Number of injections found during the search of mock data using the general template bank, "TESLA-X (main) bank" and "TESLA-X (component mass) bank" respectively.

Properties	Injected super-threshold signal
UTC time	May 10 2022 18 : 38 : 17
GPS time	1336243115.828
Distance (Mpc)	10066.811
Primary mass $m_1^{\rm det}$	$166.90 M_\odot$
Secondary mass $m_2^{\rm det}$	$117.04 M_\odot$
Dimensionless spins	$\chi_{1x} = 0.635, \chi_{1y} = 0.233, \chi_{1z} = 0.660,$ $\chi_{2x} = 0.006, \chi_{2y} = 0.087, \chi_{2z} = 0.029,$
Right ascension α	5.69
Declination δ	1.45
Inclination ι	5.99
Polarization Ψ	2.92
Waveform	IMRPhenomXPHMpseudoFourPN

Table 6.4: Summary of the properties of the injected super-threshold GW signal MS220510ae used in the simulation campaign, with all quantities given in the detector frame.

recovered by the full template bank. For clarity, we present only the Gaussian KDE corresponding to the TESLA-X main bank in the $\mathcal{M}c$ versus χeff space, and that of the TESLA-X component mass bank in the m_1 versus m_2 space, shown in Figures 6.15 and 6.16, respectively.

As in the previous case, the TESLA-X main bank and TESLA-X component mass bank are constructed by retaining all templates enclosed by the lowest density contour of the respective Gaussian KDEs. For MS220510ae, the TESLA-X main bank contains 15655 templates, while the TESLA-X component mass bank includes 21676 templates. Unlike the MS220508a case, the difference in size between the two banks is modest, with the component mass bank exceeding the main bank by only about 5000 templates.

We then apply both TESLA-X banks, together with their corresponding population models, to recover the same set of lensed injections used in the injection campaign. The recovery statistics are summarized in Table 6.5. In contrast to the results obtained for MS220508a, the TESLA-X component mass bank outperforms even the full template bank, whereas the TESLA-X main bank performs worse than the full bank. This behavior can be attributed to the fact that spin parameters are generally poorly constrained for high mass systems. Consequently, the TESLA-X main bank, which relies on constraints in the spin dimension, is more prone to missing signals, while the TESLA-X component mass bank, which imposes no spin restrictions, retains higher recovery efficiency. For completeness, we also present the sensitive range as a function of false-alarm rate for MS220510ae like signals using the full template bank, the

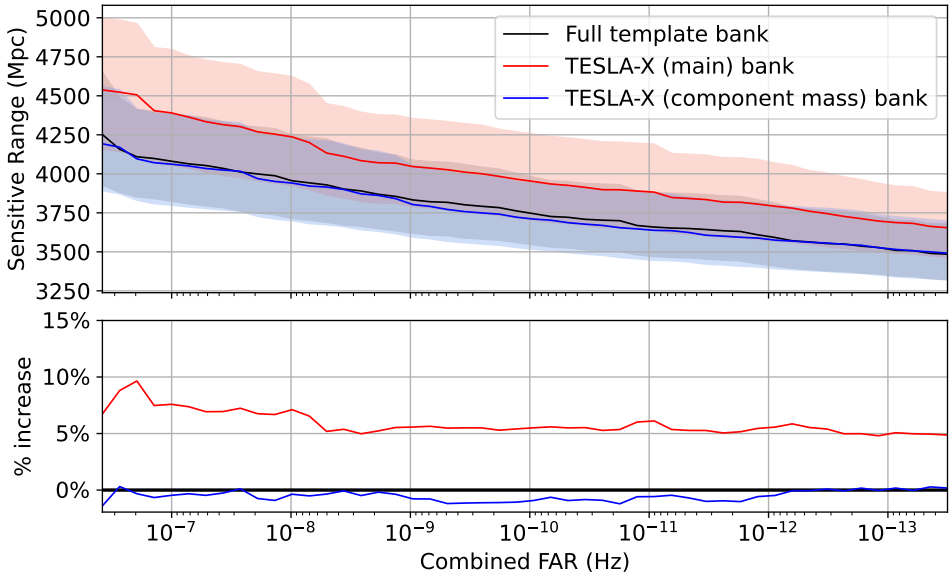

Figure 6.14: (Top panel) The sensitive range v.s. false-alarm rate for MS220508a-alike signals using the full template bank (black), "TESLA-X (main) bank" (red), and "TESLA-X (component mass) bank" (blue) respectively. The shaded band for each curve represents the corresponding 1-sigma region. (Bottom panel) The corresponding percentage changes in sensitive range v.s. false-alarm rate for the different banks. We note that while the "TESLA-X (main) bank" improves in terms of sensitivity towards MS220508a-alike (lensed) signals, "TESLA-X (component mass) bank" is performing worse than the full template bank, mainly due to the significant increase in trials factors caused by the large number of templates.

TESLA-X main bank, and the TESLA-X component mass bank in Figure 6.17.

From these two studies, we find that the choice of parameters used to construct the Gaussian KDE in the TESLA-X method should depend on the mass scale of the target event. For systems with low chirp mass, the chirp mass and effective spin provide a more suitable parameter pair for building the KDE. In contrast, for high chirp mass systems, the component masses are a more effective choice for defining the KDE. This behavior is physically expected. Chirp mass and spin parameters are typically well constrained for low mass binaries, whereas they become increasingly poorly measured for high mass systems due to the reduced number of inspiral cycles in band.

Looking ahead, several avenues can further improve the KDE construction in TESLA-X. One issue is that Gaussian KDE contours may extend beyond the physical boundaries of the original template bank, as illustrated by the yellow boundary in Figure 6.13. Parameter regions near these boundaries may therefore require additional treatment. Another promising direction is to first construct and apply a separate KDE to flatten the TESLA-X bank, for example by imposing a uniform mass distribution, before applying the Gaussian KDE that encodes the targeted population model.

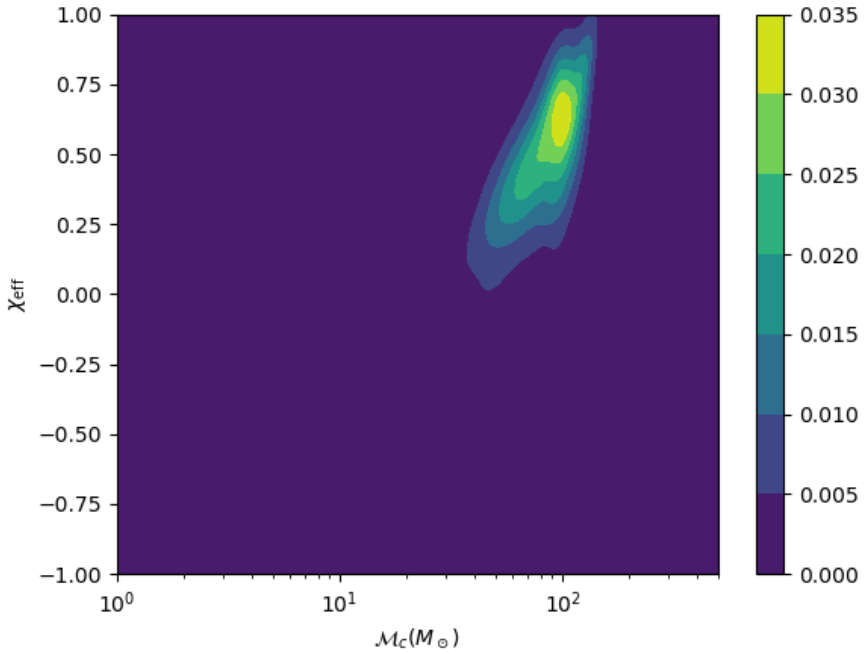

Figure 6.15: Contour map of the Gaussian kernel density estimation derived for MS220510ae in the $\mathcal{M}_c$ versus $\chi_{\rm eff}$ space, with the color scale representing the probability density estimated by the KDE.

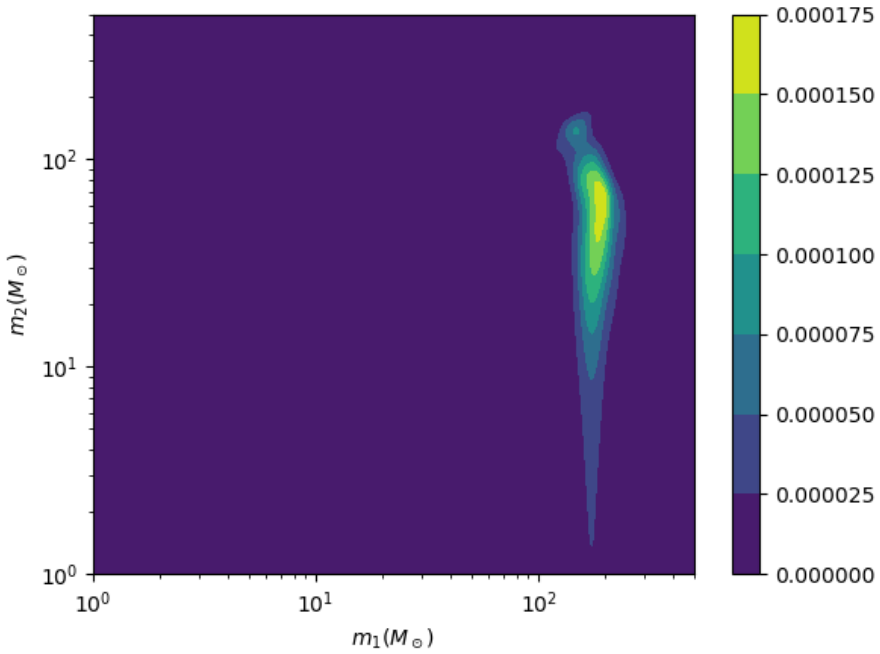

Figure 6.16: Contour map of the Gaussian kernel density estimation constructed for MS220510ae in the m_1 versus m_2 space, where the color scale indicates the probability density inferred from the KDE.

Injections	General	TESLA-X (main)	TESLA-X (component mass)
Total	8099	8099	8099
Found	5047	4715	5886
Found % change	-	−6.58%	16.6%

Table 6.5: Number of injections recovered in the mock data search for the event MS220510ae using the general template bank, the TESLA-X main bank, and the TESLA-X component mass bank, respectively.

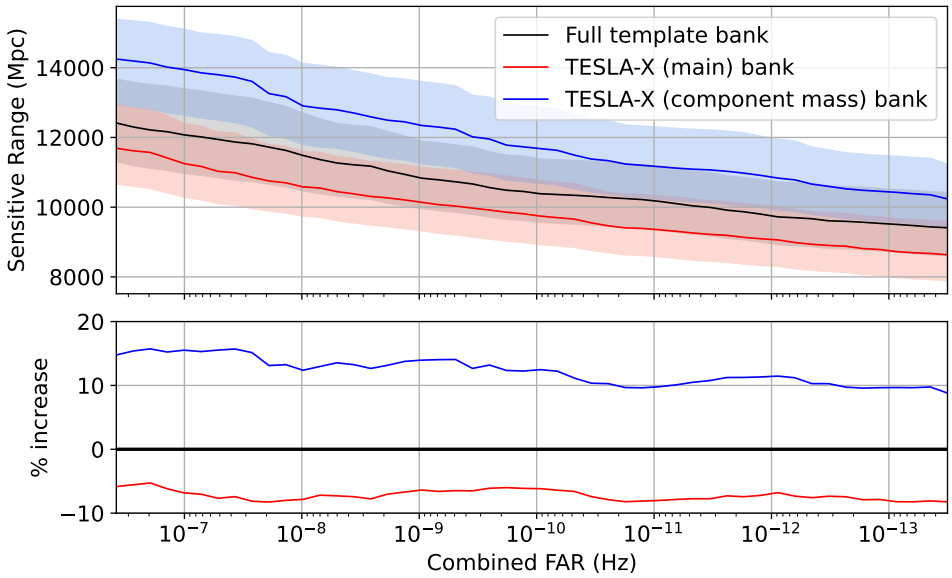

Figure 6.17: Top panel shows the sensitive range versus false-alarm rate for MS220510ae like signals obtained with the full template bank shown in black, the TESLA-X main bank shown in red, and the TESLA-X component mass bank shown in blue. The shaded region around each curve indicates the corresponding one sigma uncertainty. Bottom panel displays the fractional change in sensitive range as a function of false-alarm rate for the different template banks. Unlike the MS220508a case discussed in the main text, the TESLA-X component mass bank yields improved sensitivity to MS220510ae like lensed signals, whereas the TESLA-X main bank underperforms relative to the full template bank. This difference primarily arises because spin parameters are generally poorly constrained for systems with high chirp mass.

Bibliography

[1] Juno C. L. Chan et al. "Detectability of lensed gravitational waves in matched-filtering searches". In: *Phys. Rev. D* 111.8 (2025), p. 084019. DOI: 10.1103/PhysRevD.111.084019. arXiv: 2411.13058 [gr-qc].

[2] Juno C. L. Chan et al. "Identification and characterization of distorted gravitational waves by lensing using deep learning". In: (Nov. 2025). arXiv: 2511.07186 [gr-qc].

[3] Juno C. L. Chan et al. "Discovering gravitational waveform distortions from lensing: a deep dive into GW231123". In: (Dec. 2025). arXiv: 2512.16916 [astro-ph.CO].

[4] Alvin K. Y. Li et al. "TESLA-X: An effective method to search for sub-threshold lensed gravitational waves with a targeted population model". In: *Mon. Not. Roy. Astron. Soc.* 998 (2025), p. 1010. DOI: 10.1093/mnras/staf1259. arXiv: 2311.06416 [gr-qc].

[5] "GWTC-4.0: Searches for Gravitational-Wave Lensing Signatures". In: (Dec. 2025). arXiv: 2512.16347 [gr-qc].

[6] Juno C. L. Chan et al. "Lensing and wave optics in the strong field of a black hole". In: *Phys. Rev. D* 112.6 (2025), p. 064009. DOI: 10.1103/6h6r-46cd. arXiv: 2502.14073 [gr-qc].

[7] Mick Wright et al. "LensingFlow: An Automated Workflow for Gravitational Wave Lensing Analyses". In: (July 2025). arXiv: 2507.20256 [gr-qc].

[8] Graham P. Smith et al. "Multi-messenger gravitational lensing". In: *Phil. Trans. Roy. Soc. Lond. A* 383.2295 (2025), p. 20240134. DOI: 10.1098/rsta.2024.0134. arXiv: 2503.19973 [astro-ph.HE].

[9] Luka Vujeva et al. "Effects of galaxy cluster structure on lensed gravitational waves". In: *Phys. Rev. D* 112.6 (2025), p. 063044. DOI: 10.1103/1hmj-pxjr. arXiv: 2501.02096 [astro-ph.CO].

[10] Juno C. L. Chan, Harry Ho-Yin Ng, and Patrick Chi-Kit Cheong. "Distinct neutrino signatures of quark deconfinement in accretion-induced collapse of white dwarfs". In: *arXiv e-prints* (Dec. 2024). arXiv: 2412.10046 [hep-th].

[11] Patrick Chi-Kit Cheong et al. "General-relativistic resistive-magnetohydrodynamics simulations of self-consistent magnetized rotating neutron stars". In: *Phys. Rev. D* 111.6 (2025), p. 063030. DOI: 10.1103/PhysRevD.111.063030. arXiv: 2409.10508 [astro-ph.HE].

[12] Justin Janquart et al. "What is the nature of GW230529? An exploration of the gravitational lensing hypothesis". In: (Sept. 2024). DOI: 10.1093/mnras/staf049. arXiv: 2409.07298 [gr-qc].

[13] Rico K. L. Lo et al. “Observational Signatures of Highly Magnified Gravitational Waves from Compact Binary Coalescence”. In: *Phys. Rev. Lett.* 134.15 (2025), p. 151401. DOI: 10.1103/PhysRevLett.134.151401. arXiv: 2407.17547 [gr-qc].

[14] L. Vujeva et al. “lenscat: a public and communVujeva:2024scqity-contributed catalogue of known strong gravitationallenses”. In: *Phil. Trans. Roy. Soc. Lond. A* 383.2294 (2025), p. 20240168. DOI: 10.1098/rsta.2024.0168. arXiv: 2406.04398 [astro-ph.GA].

[15] Srashti Goyal et al. “Rapid method for preliminary identification of subthreshold strongly lensed counterparts to superthreshold gravitational-wave events”. In: *Phys. Rev. D* 109.2 (2024), p. 023028. DOI: 10.1103/PhysRevD.109.023028. arXiv: 2306.04397 [gr-qc].

[16] R. Abbott et al. “Search for Gravitational-lensing Signatures in the Full Third Observing Run of the LIGO–Virgo Network”. In: *Astrophys. J.* 970.2 (2024), p. 191. DOI: 10.3847/1538-4357/ad3e83. arXiv: 2304.08393 [gr-qc].

[17] Justin Janquart et al. “Follow-up analyses to the O3 LIGO–Virgo–KAGRA lensing searches”. In: *Mon. Not. Roy. Astron. Soc.* 526.3 (2023), pp. 3832–3860. DOI: 10.1093/mnras/stad2909. arXiv: 2306.03827 [gr-qc].

[18] Alvin K. Y. Li et al. “Targeted subthreshold search for strongly lensed gravitational-wave events”. In: *Phys. Rev. D* 107.12 (2023), p. 123014. DOI: 10.1103/PhysRevD.107.123014. arXiv: 1904.06020 [gr-qc].

[19] R. Abbott et al. “Search for subsolar-mass black hole binaries in the second part of Advanced LIGO’s and Advanced Virgo’s third observing run”. In: *Mon. Not. Roy. Astron. Soc.* 524.4 (2023). [Erratum: Mon.Not.Roy.Astron.Soc. 526, 6234 (2023)], pp. 5984–5992. DOI: 10.1093/mnras/stad588. arXiv: 2212.01477 [astro-ph.HE].

[20] R. Abbott et al. “Search for Lensing Signatures in the Gravitational-Wave Observations from the First Half of LIGO–Virgo’s Third Observing Run”. In: *Astrophys. J.* 923.1 (2021), p. 14. DOI: 10.3847/1538-4357/ac23db. arXiv: 2105.06384 [gr-qc].

[21] Michele Maggiore. *Gravitational Waves. Vol. 1: Theory and Experiments*. Oxford University Press, 2007. ISBN: 978-0-19-171766-6, 978-0-19-852074-0. DOI: 10.1093/acprof:oso/9780198570745.001.0001.

[22] Bruce Allen et al. “FINDCHIRP: An Algorithm for detection of gravitational waves from inspiraling compact binaries”. In: *Phys. Rev. D* 85 (2012), p. 122006. DOI: 10.1103/PhysRevD.85.122006. arXiv: gr-qc/0509116.

[23] Cody Messick et al. “Analysis Framework for the Prompt Discovery of Compact Binary Mergers in Gravitational-wave Data”. In: *Phys. Rev. D* 95.4 (2017), p. 042001. DOI: 10.1103/PhysRevD.95.042001. arXiv: 1604.04324 [astro-ph.IM].

[24] Eric Thrane and Colm Talbot. “An introduction to Bayesian inference in gravitational-wave astronomy: parameter estimation, model selection, and hierarchical models”. In: *Publ. Astron. Soc. Austral.* 36 (2019). [Erratum: Publ.Astron.Soc.Austral. 37, e036 (2020)], e010. DOI: 10.1017/pasa.2019.2. arXiv: 1809.02293 [astro-ph.IM].

[25] Peter Schneider, Jürgen Ehlers, and Emilio E. Falco. *Gravitational Lenses*. Astronomy and Astrophysics Library. Springer, 1992. ISBN: 978-3-540-66506-9, 978-3-662-03758-4. DOI: 10.1007/978-3-662-03758-4.

[26] Ryuichi Takahashi and Takashi Nakamura. "Wave effects in gravitational lensing of gravitational waves from chirping binaries". In: *Astrophys. J.* 595 (2003), pp. 1039–1051. DOI: 10.1086/377430. arXiv: astro-ph/0305055.

[27] P. C. Peters. "Gravitational Radiation and the Motion of Two Point Masses". In: *Phys. Rev.* 136 (1964), B1224–B1232. DOI: 10.1103/PhysRev.136.B1224.

[28] Luc Blanchet et al. "Gravitational radiation damping of compact binary systems to second post-Newtonian order". In: *Phys. Rev. Lett.* 74 (1995), pp. 3515–3518. DOI: 10.1103/PhysRevLett.74.3515. arXiv: gr-qc/9501027.

[29] P. Ajith et al. "Inspiral-merger-ringdown waveforms for black-hole binaries with non-precessing spins". In: *Phys. Rev. Lett.* 106 (2011), p. 241101. DOI: 10.1103/PhysRevLett.106.241101. arXiv: 0909.2867 [gr-qc].

[30] L. Santamaria et al. "Matching post-Newtonian and numerical relativity waveforms: systematic errors and a new phenomenological model for non-precessing black hole binaries". In: *Phys. Rev. D* 82 (2010), p. 064016. DOI: 10.1103/PhysRevD.82.064016. arXiv: 1005.3306 [gr-qc].

[31] Tito Dal Canton et al. "Real-time Search for Compact Binary Mergers in Advanced LIGO and Virgo's Third Observing Run Using PyCBC Live". In: *Astrophys. J.* 923.2 (2021), p. 254. DOI: 10.3847/1538-4357/ac2f9a. arXiv: 2008.07494 [astro-ph.HE].

[32] Qi Chu et al. "SPIIR online coherent pipeline to search for gravitational waves from compact binary coalescences". In: *Phys. Rev. D* 105.2 (2022), p. 024023. DOI: 10.1103/PhysRevD.105.024023. arXiv: 2011.06787 [gr-qc].

[33] Becca Ewing et al. "Performance of the low-latency GstLAL inspiral search towards LIGO, Virgo, and KAGRA's fourth observing run". In: *Phys. Rev. D* 109.4 (2024), p. 042008. DOI: 10.1103/PhysRevD.109.042008. arXiv: 2305.05625 [gr-qc].

[34] T. Mishra et al. "Search for binary black hole mergers in the third observing run of Advanced LIGO-Virgo using coherent WaveBurst enhanced with machine learning". In: *Phys. Rev. D* 105.8 (2022), p. 083018. DOI: 10.1103/PhysRevD.105.083018. arXiv: 2201.01495 [gr-qc].

[35] Yumeng Xu, Shubhanshu Tiwari, and Marco Drago. "PycWB: A User-friendly, Modular, and Python-based Framework for Gravitational Wave Unmodelled Search". In: (Aug. 2023). DOI: 10.1016/j.softx.2024.101639. arXiv: 2308.08639 [gr-qc].

[36] Heather Kin Yee Fong. "From Simulations to Signals: Analyzing Gravitational Waves from Compact Binary Coalescences". PhD thesis. Department of Physics, University of Toronto, July 2018.

[37] Ka Yue Alvin Li. "Probing the Higher Redshift Universe by Studying Strong Lensing of Gravitational Waves and Enhancing Search Sensitivity of the GstLAL Search Pipeline". PhD thesis. Caltech, 2024. DOI: 10.7907/hc11-h960.

[38] A. Krolak and Bernard F. Schutz. "Coalescing binaries — Probe of the universe". In: *Gen. Rel. Grav.* 19 (1987), pp. 1163–1171. DOI: 10.1007/BF00759095.

[39] R. Abbott et al. “GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run”. In: *Phys. Rev. D* 109.2 (2024), p. 022001. DOI: 10.1103/PhysRevD.109.022001. arXiv: 2108.01045 [gr-qc].

[40] Ruxandra Bondarescu et al. “Compact binaries through a lens: Silent versus detectable microlensing for the LIGO-Virgo-KAGRA gravitational wave observatories”. In: *Phys. Rev. D* 108.8 (2023), p. 084033. DOI: 10.1103/PhysRevD.108.084033. arXiv: 2211.13604 [gr-qc].

[41] Surabhi Sachdev et al. “The GstLAL Search Analysis Methods for Compact Binary Mergers in Advanced LIGO’s Second and Advanced Virgo’s First Observing Runs”. In: (Jan. 2019). arXiv: 1901.08580 [gr-qc].

[42] Kipp Cannon, Chad Hanna, and Jacob Peoples. “Likelihood-Ratio Ranking Statistic for Compact Binary Coalescence Candidates with Rate Estimation”. In: *arXiv e-prints* (Apr. 2015). arXiv: 1504.04632 [astro-ph.IM].

[43] Vaibhav Tiwari. “Estimation of the Sensitive Volume for Gravitational-wave Source Populations Using Weighted Monte Carlo Integration”. In: *Class. Quant. Grav.* 35.14 (2018), p. 145009. DOI: 10.1088/1361-6382/aac89d. arXiv: 1712.00482 [astro-ph.HE].

[44] Doğa Veske et al. “Characterizing the Observation Bias in Gravitational-wave Detections and Finding Structured Population Properties”. In: *Astrophys. J.* 922.2 (2021), p. 258. DOI: 10.3847/1538-4357/ac27ac. arXiv: 2105.13983 [gr-qc].

[45] Davide Gerosa, Geraint Pratten, and Alberto Vecchio. “Gravitational-wave selection effects using neural-network classifiers”. In: *Phys. Rev. D* 102.10 (2020), p. 103020. DOI: 10.1103/PhysRevD.102.103020. arXiv: 2007.06585 [astro-ph.HE].

[46] Jerzy Neyman and Egon Sharpe Pearson. “On the Problem of the Most Efficient Tests of Statistical Hypotheses”. In: *Phil. Trans. Roy. Soc. Lond. A* 231.694-706 (1933), pp. 289–337. DOI: 10.1098/rsta.1933.0009.

[47] Anuj Mishra et al. “Exploring the impact of microlensing on gravitational wave signals: Biases, population characteristics, and prospects for detection”. In: *Mon. Not. Roy. Astron. Soc.* 531.1 (2024), pp. 764–787. DOI: 10.1093/mnras/stae836. arXiv: 2306.11479 [astro-ph.CO].

[48] Reed Essick. “Semianalytic sensitivity estimates for catalogs of gravitational-wave transients”. In: *Phys. Rev. D* 108.4 (2023), p. 043011. DOI: 10.1103/PhysRevD.108.043011. arXiv: 2307.02765 [gr-qc].

[49] Ryan Magee et al. “Impact of selection biases on tests of general relativity with gravitational-wave inspirals”. In: *Phys. Rev. D* 109.2 (2024), p. 023014. DOI: 10.1103/PhysRevD.109.023014. arXiv: 2311.03656 [gr-qc].

[50] Antoni Ramos-Buades et al. “Impact of eccentricity on the gravitational wave searches for binary black holes: High mass case”. In: *Phys. Rev. D* 102.4 (2020), p. 043005. DOI: 10.1103/PhysRevD.102.043005. arXiv: 2005.14016 [gr-qc].

[51] Collin Capano, Yi Pan, and Alessandra Buonanno. “Impact of higher harmonics in searching for gravitational waves from nonspinning binary black holes”. In: *Phys. Rev. D* 89.10 (2014), p. 102003. DOI: 10.1103/PhysRevD.89.102003. arXiv: 1311.1286 [gr-qc].

[52] Gregory Ashton et al. “Bilby: A user-friendly Bayesian inference library for gravitational-wave astronomy”. In: *Astrophys. J. Suppl. Ser.* 241.2 (2019), p. 27. DOI: 10.3847/1538-4365/ab06fc. arXiv: 1811.02042 [astro-ph.IM].

[53] I. M. Romero-Shaw et al. “Bayesian inference for compact binary coalescences with bilby: validation and application to the first LIGO–Virgo gravitational-wave transient catalogue”. In: *Mon. Not. Roy. Astron. Soc.* 499.3 (2020), pp. 3295–3319. DOI: 10.1093/mnras/staa2850. arXiv: 2006.00714 [astro-ph.IM].

[54] R. Abbott et al. *GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — Parameter estimation data release*. Nov. 2021.

[55] A. G. Abac et al. “GWTC-4.0: An Introduction to Version 4.0 of the Gravitational-Wave Transient Catalog”. In: (Aug. 2025). arXiv: 2508.18080 [gr-qc].

[56] Stephen R. Green, Christine Simpson, and Jonathan Gair. “Gravitational-wave parameter estimation with autoregressive neural network flows”. In: *Phys. Rev. D* 102.10 (2020), p. 104057. DOI: 10.1103/PhysRevD.102.104057. arXiv: 2002.07656 [astro-ph.IM].

[57] Stephen R. Green and Jonathan Gair. “Complete parameter inference for GW150914 using deep learning”. In: *Mach. Learn. Sci. Tech.* 2.3 (2021), 03LT01. DOI: 10.1088/2632-2153/abfaed. arXiv: 2008.03312 [astro-ph.IM].

[58] Maximilian Dax et al. “Real-Time Gravitational Wave Science with Neural Posterior Estimation”. In: *Phys. Rev. Lett.* 127.24 (2021), p. 241103. DOI: 10.1103/PhysRevLett.127.241103. arXiv: 2106.12594 [gr-qc].

[59] Maximilian Dax et al. “Group equivariant neural posterior estimation”. In: (Nov. 2021). arXiv: 2111.13139 [cs.LG].

[60] Jonas Wildberger et al. “Adapting to noise distribution shifts in flow-based gravitational-wave inference”. In: *Phys. Rev. D* 107.8 (2023), p. 084046. DOI: 10.1103/PhysRevD.107.084046. arXiv: 2211.08801 [gr-qc].

[61] Maximilian Dax et al. “Neural Importance Sampling for Rapid and Reliable Gravitational-Wave Inference”. In: *Phys. Rev. Lett.* 130.17 (2023), p. 171403. DOI: 10.1103/PhysRevLett.130.171403. arXiv: 2210.05686 [gr-qc].

[62] Nihar Gupte et al. “Evidence for eccentricity in the population of binary black holes observed by LIGO-Virgo-KAGRA”. In: (Apr. 2024). arXiv: 2404.14286 [gr-qc].

[63] Anna Moreso Serra and Oleg Bulashenko. “Probing binary lens caustics with gravitational waves: A uniform approximation approach”. In: *Phys. Rev. D* 112.2 (2025), p. 024041. DOI: 10.1103/qkvt-yp6r. arXiv: 2504.10128 [gr-qc].

[64] Jose María Ezquiaga, Rico K. L. Lo, and Luka Vujeva. “Diffraction around caustics in gravitational wave lensing”. In: *Phys. Rev. D* 112.4 (2025), p. 043544. DOI: 10.1103/tkk2-x9st. arXiv: 2503.22648 [gr-qc].

[65] Shuji Deguchi and William D. Watson. “Wave effects in gravitational lensing of electromagnetic radiation”. In: *Phys. Rev. D* 34 (1986), pp. 1708–1718. DOI: 10.1103/PhysRevD.34.1708.

[66] Takahiro T. Nakamura. "Gravitational lensing of gravitational waves from inspiraling binaries by a point mass lens". In: *Phys. Rev. Lett.* 80 (1998), pp. 1138–1141. DOI: 10.1103/PhysRevLett.80.1138.

[67] Calvin Leung et al. "Wave Mechanics, Interference, and Decoherence in Strong Gravitational Lensing". In: (Apr. 2023). arXiv: 2304.01202 [astro-ph.HE].

[68] Anna Liu et al. "Millilensing induced systematic biases in parameterized tests of General Relativity". In: (Oct. 2024). arXiv: 2410.21738 [gr-qc].

[69] Sunghoon Jung and Chang Sub Shin. "Gravitational-Wave Fringes at LIGO: Detecting Compact Dark Matter by Gravitational Lensing". In: *Phys. Rev. Lett.* 122.4 (2019), p. 041103. DOI: 10.1103/PhysRevLett.122.041103. arXiv: 1712.01396 [astro-ph.CO].

[70] Kwun-Hang Lai et al. "Discovering intermediate-mass black hole lenses through gravitational wave lensing". In: *Phys. Rev. D* 98.8 (2018), p. 083005. DOI: 10.1103/PhysRevD.98.083005. arXiv: 1801.07840 [gr-qc].

[71] Pierre Christian, Salvatore Vitale, and Abraham Loeb. "Detecting Stellar Lensing of Gravitational Waves with Ground-Based Observatories". In: *Phys. Rev. D* 98.10 (2018), p. 103022. DOI: 10.1103/PhysRevD.98.103022. arXiv: 1802.02586 [astro-ph.HE].

[72] Liang Dai et al. "Detecting Lensing-Induced Diffraction in Astrophysical Gravitational Waves". In: *Phys. Rev. D* 98.10 (2018), p. 104029. DOI: 10.1103/PhysRevD.98.104029. arXiv: 1810.00003 [gr-qc].

[73] J.M. Diego et al. "Observational signatures of microlensing in gravitational waves at LIGO/Virgo frequencies". In: *Astron. Astrophys.* 627 (2019), A130. DOI: 10.1051/0004-6361/201935490. arXiv: 1903.04513 [astro-ph.CO].

[74] Jose M. Diego. "Constraining the abundance of primordial black holes with gravitational lensing of gravitational waves at LIGO frequencies". In: *Phys. Rev. D* 101.12 (2020), p. 123512. DOI: 10.1103/PhysRevD.101.123512. arXiv: 1911.05736 [astro-ph.CO].

[75] Kai Liao, Shuxun Tian, and Xuheng Ding. "Probing compact dark matter with gravitational wave fringes detected by the Einstein Telescope". In: *Mon. Not. Roy. Astron. Soc.* 495.2 (2020), pp. 2002–2006. DOI: 10.1093/mnras/staa1388. arXiv: 2001.07891 [astro-ph.CO].

[76] Anuj Mishra et al. "Gravitational lensing of gravitational waves: effect of microlens population in lensing galaxies". In: *Mon. Not. Roy. Astron. Soc.* 508.4 (2021), pp. 4869–4886. DOI: 10.1093/mnras/stab2875. arXiv: 2102.03946 [astro-ph.CO].

[77] Anna Liu et al. "Exploring the hidden Universe: a novel phenomenological approach for recovering arbitrary gravitational-wave millilensing configurations". In: *Mon. Not. Roy. Astron. Soc.* 525.3 (2023), pp. 4149–4160. DOI: 10.1093/mnras/stad1302. arXiv: 2302.09870 [gr-qc].

[78] Aniruddha Chakraborty and Suvodip Mukherjee. "μ-GLANCE: A Novel Technique to Detect Chromatically and Achromatically Lensed Gravitational Wave Signals". In: (Oct. 2024). arXiv: 2410.06995 [gr-qc].

[79] Malcolm Fairbairn, Juan Urrutia, and Ville Vaskonen. "Microlensing of gravitational waves by dark matter structures". In: *Journal of Cosmology and Astroparticle Physics* 2023.07 (2023), p. 007.

[80] Giovanni Tambalo et al. “Gravitational wave lensing as a probe of halo properties and dark matter”. In: *Phys. Rev. D* 108.10 (2023), p. 103529. DOI: 10.1103/PhysRevD.108.103529. arXiv: 2212.11960 [astro-ph.CO].

[81] Mesut Çalışkan et al. “Observability of lensing of gravitational waves from massive black hole binaries with LISA”. In: *Phys. Rev. D* 107.4 (2023), p. 043029. DOI: 10.1103/PhysRevD.107.043029. arXiv: 2206.02803 [astro-ph.CO].

[82] Stefano Savastano et al. “Weakly lensed gravitational waves: Probing cosmic structures with wave-optics features”. In: *Phys. Rev. D* 108.10 (2023), p. 103532. DOI: 10.1103/PhysRevD.108.103532. arXiv: 2306.05282 [gr-qc].

[83] M. V. S. Saketh, Rajes Ghosh, and Anuj Mishra. “Strong-field Gravitational Wave Lensing in the Kerr Background”. In: (Nov. 2025). arXiv: 2511.23110 [gr-qc].

[84] Zhao Li and Wen Zhao. “Rigorous calculation of scalar scattering in the Schwarzschild background: The convergence of the partial-wave series and the Poisson spot”. In: *Phys. Rev. D* 112.8 (2025), p. 083030. DOI: 10.1103/xsxj-9vdw. arXiv: 2508.17253 [gr-qc].

[85] Mariana Carrillo Gonzalez et al. “A scattering perspective on gravitational lensing”. In: (Nov. 2025). arXiv: 2511.15797 [hep-th].

[86] Zhao Li, Shaoqi Hou, and Wen Zhao. “Gravitational Lensing of Gravitational Waves: Spin-wave Optics through Black Hole Scattering”. In: (Dec. 2025). arXiv: 2512.23933 [gr-qc].

[87] Gerasimos Kouniatalis, Arthur G. Suvorov, and Kyriakos Destounis. “Lensing by black holes within astrophysical environments”. In: *Phys. Rev. D* 112.12 (2025), p. 124062. DOI: 10.1103/tn1g-7tfj. arXiv: 2508.19333 [gr-qc].

[88] Paolo Cremonese, Jose María Ezquiaga, and Vincenzo Salzano. “Breaking the mass-sheet degeneracy with gravitational wave interference in lensed events”. In: *Phys. Rev. D* 104.2 (2021), p. 023503. DOI: 10.1103/PhysRevD.104.023503. arXiv: 2104.07055 [astro-ph.CO].

[89] Anna Balaudo et al. “Prospects of testing late-time cosmology with weak lensing of gravitational waves and galaxy surveys”. In: *JCAP* 06 (2023), p. 050. DOI: 10.1088/1475-7516/2023/06/050. arXiv: 2210.06398 [astro-ph.CO].

[90] Anson Chen et al. “Invariance transformations in wave-optics lensing: implications for gravitational-wave astrophysics and cosmology”. In: (Aug. 2024). arXiv: 2408.03856 [astro-ph.CO].

[91] Zhoujian Cao, Li-Fang Li, and Yan Wang. “Gravitational lensing effects on parameter estimation in gravitational wave detection with advanced detectors”. In: *Phys. Rev. D* 90.6 (2014), p. 062003. DOI: 10.1103/PhysRevD.90.062003.

[92] Jose María Ezquiaga, Wayne Hu, and Macarena Lagos. “Apparent Superluminality of Lensed Gravitational Waves”. In: *Phys. Rev. D* 102.2 (2020), p. 023531. DOI: 10.1103/PhysRevD.102.023531. arXiv: 2005.10702 [astro-ph.CO].

[93] Anuradha Gupta et al. “Possible Causes of False General Relativity Violations in Gravitational Wave Observations”. In: (May 2024). arXiv: 2405.02197 [gr-qc].

[94] Yun Wang, Albert Stebbins, and Edwin L. Turner. "Gravitational lensing of gravitational waves from merging neutron star binaries". In: *Phys. Rev. Lett.* 77 (1996), pp. 2875–2878. DOI: 10.1103/PhysRevLett.77.2875. arXiv: astro-ph/9605140.

[95] Liang Dai and Tejaswi Venumadhav. "On the waveforms of gravitationally lensed gravitational waves". In: *arXiv e-prints*, arXiv:1702.04724 (Feb. 2017), arXiv:1702.04724. DOI: 10.48550/arXiv.1702.04724. arXiv: 1702.04724 [gr-qc].

[96] J. M. Ezquiaga et al. "Phase effects from strong gravitational lensing of gravitational waves". In: *Phys. Rev. D* 103 (2021), p. 6. DOI: 10.1103/PhysRevD.103.064047. arXiv: 2008.12814 [gr-qc].

[97] Rico K. L. Lo and Ignacio Magana Hernandez. "Bayesian statistical framework for identifying strongly lensed gravitational-wave signals". In: *Phys. Rev. D* 107.12 (2023), p. 123015. DOI: 10.1103/PhysRevD.107.123015. arXiv: 2104.09339 [gr-qc].

[98] Yijun Wang et al. "Identifying Type II Strongly Lensed Gravitational-Wave Images in Third-Generation Gravitational-Wave Detectors". In: *Phys. Rev. D* 103.10 (2021), p. 104055. DOI: 10.1103/PhysRevD.103.104055. arXiv: 2101.08264 [gr-qc].

[99] Srashti Goyal et al. "Testing the nature of gravitational-wave polarizations using strongly lensed signals". In: *Phys. Rev. D* 103.2 (2021), p. 024038. DOI: 10.1103/PhysRevD.103.024038. arXiv: 2008.07060 [gr-qc].

[100] Andreas Finke et al. "Probing modified gravitational wave propagation with strongly lensed coalescing binaries". In: *Phys. Rev. D* 104.8 (2021), p. 084057. DOI: 10.1103/PhysRevD.104.084057. arXiv: 2107.05046 [gr-qc].

[101] Srashti Goyal et al. "Probing lens-induced gravitational-wave birefringence as a test of general relativity". In: *Phys. Rev. D* 108.2 (2023), p. 024052. DOI: 10.1103/PhysRevD.108.024052. arXiv: 2301.04826 [gr-qc].

[102] Harsh Narola et al. "How well can modified gravitational wave propagation be constrained with strong lensing?" In: *Phys. Rev. D* 109.8 (2024), p. 084064. DOI: 10.1103/PhysRevD.109.084064. arXiv: 2308.01709 [gr-qc].

[103] Otto A. Hannuksela et al. "Localizing merging black holes with sub-arcsecond precision using gravitational-wave lensing". In: *Mon. Not. Roy. Astron. Soc.* 498.3 (2020), pp. 3395–3402. DOI: 10.1093/mnras/staa2577. arXiv: 2004.13811 [astro-ph.HE].

[104] Laura Elina Uronen et al. "Finding black holes: an unconventional multi-messenger". In: *Phil. Trans. Roy. Soc. Lond. A* 383.2294 (2025), p. 20240152. DOI: 10.1098/rsta.2024.0152. arXiv: 2406.14257 [astro-ph.HE].

[105] Sourabh Magare et al. "Gear Up for the Action Replay: Leveraging Lensing for Enhanced Gravitational-wave Early Warning". In: *Astrophys. J. Lett.* 955.2 (2023), p. L31. DOI: 10.3847/2041-8213/acf668. arXiv: 2302.02916 [astro-ph.HE].

[106] Kai Liao et al. "Precision cosmology from future lensed gravitational wave and electromagnetic signals". In: *Nature Commun.* 8.1 (2017). [Erratum: Nature Commun. 8, 2136 (2017)], p. 1148. DOI: 10.1038/s41467-017-01152-9. arXiv: 1703.04151 [astro-ph.CO].

[107] Zhiwei Chen et al. “Enhanced Localization of Dark Lensed Gravitational-wave Events Enables Host Galaxy Identification and Precise Cosmological Inference”. In: *Astrophys. J. Lett.* 993.2 (2025), p. L57. DOI: 10.3847/2041-8213/ae1226. arXiv: 2510.12470 [astro-ph.CO].

[108] Jason S. C. Poon et al. “Galaxy lens reconstruction based on strongly lensed gravitational waves: similarity transformation degeneracy and mass-sheet degeneracy”. In: *Mon. Not. Roy. Astron. Soc.* 536.3 (2024), pp. 2212–2233. DOI: 10.1093/mnras/stae2660. arXiv: 2406.06463 [astro-ph.HE].

[109] Thomas C. K. Ng, Stefano Rinaldi, and Otto A. Hannuksela. “Inferring cosmology from gravitational waves using non-parametric detector-frame mass distribution”. In: (Oct. 2024). arXiv: 2410.23541 [gr-qc].

[110] Bin Liu, Zhengxiang Li, and Zong-Hong Zhu. “Complementary constraints on dark energy equation of state from strongly lensed gravitational wave”. In: *Mon. Not. Roy. Astron. Soc.* 487.2 (2019), pp. 1980–1985. DOI: 10.1093/mnras/stz1179. arXiv: 1904.11751 [astro-ph.CO].

[111] Peter T. H. Pang et al. “Lensed or not lensed: determining lensing magnifications for binary neutron star mergers from a single detection”. In: *Mon. Not. Roy. Astron. Soc.* 495.4 (2020), pp. 3740–3750. DOI: 10.1093/mnras/staa1430. arXiv: 2002.04893 [astro-ph.HE].

[112] Souvik Jana et al. “Cosmography Using Strongly Lensed Gravitational Waves from Binary Black Holes”. In: *Phys. Rev. Lett.* 130.26 (2023), p. 261401. DOI: 10.1103/PhysRevLett.130.261401. arXiv: 2211.12212 [astro-ph.CO].

[113] Hsin-Yu Chen, Jose María Ezquiaga, and Ish Gupta. “Cosmography with next-generation gravitational wave detectors”. In: *Class. Quant. Grav.* 41.12 (2024), p. 125004. DOI: 10.1088/1361-6382/ad424f. arXiv: 2402.03120 [gr-qc].

[114] Koustav N. Maity et al. “Strong lensing cosmography using binary-black-hole mergers: Prospects for the near future”. In: (Dec. 2025). arXiv: 2512.15168 [gr-qc].

[115] Amanda Farah et al. “Gravitational lensing rarely produces high-mass outliers to the compact binary population”. In: (July 2025). arXiv: 2507.07964 [astro-ph.HE].

[116] Suvodip Mukherjee, Benjamin D. Wandelt, and Joseph Silk. “Probing the theory of gravity with gravitational lensing of gravitational waves and galaxy surveys”. In: *Mon. Not. Roy. Astron. Soc.* 494.2 (2020), pp. 1956–1970. DOI: 10.1093/mnras/staa827. arXiv: 1908.08951 [astro-ph.CO].

[117] Luka Vujeva et al. “Dark Matter Subhalos and Higher Order Catastrophes in Gravitational Wave Lensing”. In: (Oct. 2025). arXiv: 2510.14953 [astro-ph.CO].

[118] Ken K. Y. Ng et al. “Precise LIGO Lensing Rate Predictions for Binary Black Holes”. In: *Phys. Rev. D* 97.2 (2018), p. 023012. DOI: 10.1103/PhysRevD.97.023012. arXiv: 1703.06319 [astro-ph.CO].

[119] Masamune Oguri. “Effect of gravitational lensing on the distribution of gravitational waves from distant binary black hole mergers”. In: *Mon. Not. Roy. Astron. Soc.* 480.3 (2018), pp. 3842–3855. DOI: 10.1093/mnras/sty2145. arXiv: 1807.02584 [astro-ph.CO].

[120] A. Renske A. C. Wierda et al. “Beyond the Detector Horizon: Forecasting Gravitational-Wave Strong Lensing”. In: *Astrophys. J.* 921.2 (2021), p. 154. DOI: 10.3847/1538-4357/ac1bb4. arXiv: 2106.06303 [astro-ph.HE].

[121] Fei Xu, Jose Maria Ezquiaga, and Daniel E. Holz. “Please Repeat: Strong Lensing of Gravitational Waves as a Probe of Compact Binary and Galaxy Populations”. In: *Astrophys. J.* 929.1 (2022), p. 9. DOI: 10.3847/1538-4357/ac58f8. arXiv: 2105.14390 [astro-ph.CO].

[122] Anupreeta More and Surhud More. “Improved statistic to identify strongly lensed gravitational wave events”. In: *Mon. Not. Roy. Astron. Soc.* 515.1 (2022), pp. 1044–1051. DOI: 10.1093/mnras/stac1704. arXiv: 2111.03091 [astro-ph.CO].

[123] Graham P. Smith et al. “Discovering gravitationally lensed gravitational waves: predicted rates, candidate selection, and localization with the Vera Rubin Observatory”. In: *Mon. Not. Roy. Astron. Soc.* 520.1 (2023), pp. 702–721. DOI: 10.1093/mnras/stad140. arXiv: 2204.12977 [astro-ph.HE].

[124] Hemantakumar Phurailatpam et al. “ler : LVK (LIGO-Virgo-KAGRA collaboration) event (compact-binary mergers) rate calculator and simulator”. In: (July 2024). arXiv: 2407.07526 [astro-ph.IM].

[125] J. Aasi et al. “Advanced LIGO”. In: *Class. Quant. Grav.* 32 (2015), p. 074001. DOI: 10.1088/0264-9381/32/7/074001. arXiv: 1411.4547 [gr-qc].

[126] F. Acernese et al. “Advanced Virgo: a second-generation interferometric gravitational wave detector”. In: *Class. Quant. Grav.* 32.2 (2015), p. 024001. DOI: 10.1088/0264-9381/32/2/024001. arXiv: 1408.3978 [gr-qc].

[127] O. A. Hannuksela et al. “Search for gravitational lensing signatures in LIGO-Virgo binary black hole events”. In: *Astrophys. J. Lett.* 874.1 (2019), p. L2. DOI: 10.3847/2041-8213/ab0c0f. arXiv: 1901.02674 [gr-qc].

[128] Connor McIsaac et al. “Search for strongly lensed counterpart images of binary black hole mergers in the first two LIGO observing runs”. In: *Phys. Rev. D* 102.8 (2020), p. 084031. DOI: 10.1103/PhysRevD.102.084031. arXiv: 1912.05389 [gr-qc].

[129] Xiaoshu Liu, Ignacio Magana Hernandez, and Jolien Creighton. “Identifying strong gravitational-wave lensing during the second observing run of Advanced LIGO and Advanced Virgo”. In: *Astrophys. J.* 908.1 (2021), p. 97. DOI: 10.3847/1538-4357/abd7eb. arXiv: 2009.06539 [astro-ph.HE].

[130] Liang Dai et al. “Search for Lensed Gravitational Waves Including Morse Phase Information: An Intriguing Candidate in O2”. In: *arXiv e-prints*, arXiv:2007.12709 (July 2020), arXiv:2007.12709. DOI: 10.48550/arXiv.2007.12709. arXiv: 2007.12709 [astro-ph.HE].

[131] A. G. Abac et al. “GW231123: a Binary Black Hole Merger with Total Mass 190-265 $M_\odot$”. In: (July 2025). arXiv: 2507.08219 [astro-ph.HE].

[132] Srashti Goyal, Hector Villarrubia-Rojo, and Miguel Zumalacarregui. “Across the Universe: GW231123 as a magnified and diffracted black hole merger”. In: (Dec. 2025). arXiv: 2512.17631 [astro-ph.GA].

[133] Aniruddha Chakraborty and Suvodip Mukherjee. “The First Model-Independent Upper Bound on Micro-lensing Signature of the Highest Mass Binary Black Hole Event GW231123”. In: (Dec. 2025). arXiv: 2512.19077 [gr-qc].

[134] Xikai Shan, Huan Yang, and Shude Mao. “GW231123: A Case for Binary Microlensing in a Strong Lensing Field”. In: (Dec. 2025). arXiv: 2512.19118 [astro-ph.GA].

[135] Gregory Ashton and Colm Talbot. “B ilby-MCMC: an MCMC sampler for gravitational-wave inference”. In: *Mon. Not. Roy. Astron. Soc.* 507.2 (2021), pp. 2037–2051. DOI: 10.1093/mnras/stab2236. arXiv: 2106.08730 [gr-qc].

[136] Mick Wright and Martin Hendry. “Gravelamps: Gravitational Wave Lensing Mass Profile Model Selection”. In: *The Astrophysical Journal* 935.2, 68 (Aug. 2022), p. 68. DOI: 10.3847/1538-4357/ac7ec2. arXiv: 2112.07012 [astro-ph.HE].

[137] Mesut Çalışkan et al. “Lensing or luck? False alarm probabilities for gravitational lensing of gravitational waves”. In: *Phys. Rev. D* 107.6 (2023), p. 063023. DOI: 10.1103/PhysRevD.107.063023. arXiv: 2201.04619 [astro-ph.CO].

[138] David Keitel. “False positives for gravitational lensing: the gravitational-wave perspective”. In: *Phil. Trans. Roy. Soc. Lond. A* 383.2295 (2025), p. 20240128. DOI: 10.1098/rsta.2024.0128. arXiv: 2407.12974 [gr-qc].

[139] Ángel Garrón and David Keitel. “Waveform systematics in identifying strongly gravitationally lensed gravitational waves: posterior overlap method”. In: *Class. Quant. Grav.* 41.1 (2024), p. 015005. DOI: 10.1088/1361-6382/ad0b9b. arXiv: 2306.12908 [gr-qc].

[140] A. Barsode, K. N. Maity, and P. Ajith. “Lensing, not luck! Detection prospects of strongly lensed gravitational waves”. In: (Oct. 2025). arXiv: 2510.23238 [gr-qc].

[141] Otto A. Hannuksela et al. “Strong gravitational-wave lensing posterior odds”. In: (Oct. 2025). arXiv: 2510.15463 [astro-ph.HE].

[142] Ilya Mandel, Will M. Farr, and Jonathan R. Gair. “Extracting distribution parameters from multiple uncertain observations with selection biases”. In: *Mon. Not. Roy. Astron. Soc.* 486.1 (2019), pp. 1086–1093. DOI: 10.1093/mnras/stz896. arXiv: 1809.02063 [physics.data-an].

[143] Stephen Privitera et al. “Improving the sensitivity of a search for coalescing binary black holes with nonprecessing spins in gravitational wave data”. In: *Phys. Rev. D* 89.2 (2014), p. 024003. DOI: 10.1103/PhysRevD.89.024003. arXiv: 1310.5633 [gr-qc].

[144] Alejandro Bohé et al. “Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors”. In: *Phys. Rev. D* 95.4 (2017), p. 044028. DOI: 10.1103/PhysRevD.95.044028. arXiv: 1611.03703 [gr-qc].

[145] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. *LVK Algorithm Library - LALSuite*. Free software (GPL). 2018. DOI: 10.7935/GT1W-FZ16.

[146] A. G. Abac et al. “GW250114: Testing Hawking’s Area Law and the Kerr Nature of Black Holes”. In: *Phys. Rev. Lett.* 135.11 (2025), p. 111403. DOI: 10.1103/kw5g-d732. arXiv: 2509.08054 [gr-qc].

[147] Kipp Cannon et al. “GstLAL: A software framework for gravitational wave discovery”. In: *SoftwareX* 14, 100680 (June 2021), p. 100680. DOI: 10.1016/j.softx.2021.100680. arXiv: 2010.05082 [astro-ph.IM].

[148] Leo Tsukada et al. “Improved ranking statistics of the GstLAL inspiral search for compact binary coalescences”. In: *Phys. Rev. D* 108.4 (2023), p. 043004. DOI: 10.1103/PhysRevD.108.043004. arXiv: 2305.06286 [astro-ph.IM].

[149] Samantha A. Usman et al. “The PyCBC search for gravitational waves from compact binary coalescence”. In: *Class. Quant. Grav.* 33.21 (2016), p. 215004. DOI: 10.1088/0264-9381/33/21/215004. arXiv: 1508.02357 [gr-qc].

[150] F. Aubin et al. “The MBTA pipeline for detecting compact binary coalescences in the third LIGO–Virgo observing run”. In: *Class. Quant. Grav.* 38.9 (2021), p. 095004. DOI: 10.1088/1361-6382/abe913. arXiv: 2012.11512 [gr-qc].

[151] Christopher Alléné et al. “The MBTA pipeline for detecting compact binary coalescences in the fourth LIGO-Virgo-KAGRA observing run”. In: *Class. Quant. Grav.* 42.10 (2025), p. 105009. DOI: 10.1088/1361-6382/add234. arXiv: 2501.04598 [gr-qc].

[152] Digvijay Wadekar et al. “New search pipeline for gravitational waves with higher-order modes using mode-by-mode filtering”. In: *Phys. Rev. D* 110.4 (2024), p. 044063. DOI: 10.1103/PhysRevD.110.044063. arXiv: 2405.17400 [gr-qc].

[153] Marco Drago et al. “coherent WaveBurst, a pipeline for unmodeled gravitational-wave data analysis”. In: *SoftwareX* 14, 100678 (June 2021), p. 100678. DOI: 10.1016/j.softx.2021.100678. arXiv: 2006.12604 [gr-qc].

[154] Ashish Kumar Meena et al. “Gravitational lensing of gravitational waves: Probability of microlensing in galaxy-scale lens population”. In: *Mon. Not. Roy. Astron. Soc.* 517.1 (2022), pp. 872–884. DOI: 10.1093/mnras/stac2721. arXiv: 2205.05409 [astro-ph.GA].

[155] Masamune Oguri and Ryuichi Takahashi. “Probing Dark Low-mass Halos and Primordial Black Holes with Frequency-dependent Gravitational Lensing Dispersions of Gravitational Waves”. In: *Astrophys. J.* 901.1 (2020), p. 58. DOI: 10.3847/1538-4357/abafab. arXiv: 2007.01936 [astro-ph.CO].

[156] Juan Urrutia and Ville Vaskonen. “Lensing of gravitational waves as a probe of compact dark matter”. In: *Mon. Not. Roy. Astron. Soc.* 509.1 (2021), pp. 1358–1365. DOI: 10.1093/mnras/stab3118. arXiv: 2109.03213 [astro-ph.CO].

[157] Huan Zhou et al. “Constraints on compact dark matter from lensing of gravitational waves for the third-generation gravitational wave detector”. In: *Mon. Not. Roy. Astron. Soc.* 518.1 (2022), pp. 149–156. DOI: 10.1093/mnras/stac2944. arXiv: 2206.13128 [astro-ph.CO].

[158] Bo Liang and He Wang. “Recent Advances in Simulation-based Inference for Gravitational Wave Data Analysis”. In: (July 2025). DOI: 10.61977/ati2025020. arXiv: 2507.11192 [gr-qc].

[159] Eungwang Seo et al. “Residual Test to Search for Microlensing Signatures in Strongly Lensed Gravitational Wave Signals”. In: *Astrophys. J.* 988.2 (2025), p. 159. DOI: 10.3847/1538-4357/ade4bf. arXiv: 2503.02186 [gr-qc].

[160] Zheng Qin et al. “Parameter inference of microlensed gravitational waves using neural spline flows”. In: (May 2025). arXiv: 2505.20996 [gr-qc].

[161] Alvin J. K. Chua and Michele Vallisneri. “Learning Bayesian posteriors with neural networks for gravitational-wave inference”. In: *Phys. Rev. Lett.* 124.4 (2020), p. 041102. DOI: 10.1103/PhysRevLett.124.041102. arXiv: 1909.05966 [gr-qc].

[162] Hunter Gabbard et al. “Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy”. In: *Nature Phys.* 18.1 (2022), pp. 112–117. DOI: 10.1038/s41567-021-01425-7. arXiv: 1909.06296 [astro-ph.IM].

[163] Elena Cuoco et al. “Enhancing Gravitational-Wave Science with Machine Learning”. In: *Mach. Learn. Sci. Tech.* 2.1 (2021), p. 011002. DOI: 10.1088/2632-2153/abb93a. arXiv: 2005.03745 [astro-ph.HE].

[164] Jurriaan Langendorff et al. “Normalizing Flows as an Avenue to Studying Overlapping Gravitational Wave Signals”. In: *Phys. Rev. Lett.* 130.17 (2023), p. 171402. DOI: 10.1103/PhysRevLett.130.171402. arXiv: 2211.15097 [gr-qc].

[165] Tianyu Zhao et al. “Dawning of a new era in gravitational wave data analysis: Unveiling cosmic mysteries via artificial intelligence — A systematic review”. In: *Front. Phys. (Beijing)* 20.4 (2025), p. 45301. DOI: 10.15302/frontphys.2025.045301. arXiv: 2311.15585 [gr-qc].

[166] Uddipta Bhardwaj et al. “Sequential simulation-based inference for gravitational wave signals”. In: *Phys. Rev. D* 108.4 (2023), p. 042004. DOI: 10.1103/PhysRevD.108.042004. arXiv: 2304.02035 [gr-qc].

[167] Elena Cuoco et al. “Applications of machine learning in gravitational-wave research with current interferometric detectors”. In: *Living Rev. Rel.* 28.1 (2025), p. 2. DOI: 10.1007/s41114-024-00055-8. arXiv: 2412.15046 [gr-qc].

[168] Qian Hu et al. “Decoding Long-duration Gravitational Waves from Binary Neutron Stars with Machine Learning: Parameter Estimation and Equations of State”. In: (Dec. 2024). arXiv: 2412.03454 [gr-qc].

[169] Alicja Polanska et al. “Accelerated Bayesian parameter estimation and model selection for gravitational waves with normalizing flows”. In: *38th conference on Neural Information Processing Systems*. Oct. 2024. arXiv: 2410.21076 [astro-ph.IM].

[170] Nikolaos Stergioulas. “Machine Learning Applications in Gravitational Wave Astronomy”. In: Jan. 2024. arXiv: 2401.07406 [gr-qc].

[171] Lucia Papalini et al. “Can Transformers help us perform parameter estimation of overlapping signals in gravitational wave detectors?” In: *Class. Quant. Grav.* 42.18 (2025), p. 185012. DOI: 10.1088/1361-6382/adfd33. arXiv: 2505.02773 [gr-qc].

[172] Filippo Santoliquido et al. “Fast and accurate parameter estimation of high-redshift sources with the Einstein Telescope”. In: (Apr. 2025). arXiv: 2504.21087 [astro-ph.HE].

[173] Martin Benedikt and Ippocratis D. Saltas. “GraviBERT: Transformer-based inference for gravitational-wave time series”. In: (Dec. 2025). arXiv: 2512.21390 [gr-qc].

[174] Srashti Goyal et al. “Rapid identification of strongly lensed gravitational-wave events with machine learning”. In: *Phys. Rev. D* 104.12 (2021), p. 124057. DOI: 10.1103/PhysRevD.104.124057. arXiv: 2106.12466 [gr-qc].

[175] Sourabh Magare, Anupreeta More, and Sunil Choudhary. “SLICK: Strong Lensing Identification of Candidates Kindred in gravitational wave data”. In: *Mon. Not. Roy. Astron. Soc.* 535.1 (2024), pp. 990–999. DOI: 10.1093/mnras/stae2408. arXiv: 2403.02994 [astro-ph.HE].

[176] Kyungmin Kim et al. “Identification of Lensed Gravitational Waves with Deep Learning”. In: *Astrophys. J.* 915.2 (2021), p. 119. DOI: 10.3847/1538-4357/ac0143. arXiv: 2010.12093 [gr-qc].

[177] Roberto Bada-Nerin et al. “Parameter estimation of microlensed gravitational waves with conditional variational autoencoders”. In: *Phys. Rev. D* 111.8 (2025), p. 084067. DOI: 10.1103/PhysRevD.111.084067. arXiv: 2412.00566 [gr-qc].

[178] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. “The frontier of simulation-based inference”. In: *Proc. Nat. Acad. Sci.* 117.48 (2020), pp. 30055–30062. DOI: 10.1073/pnas.1912789117. arXiv: 1911.01429 [stat.ML].

[179] Qian Hu et al. “GW231123: Overlapping Gravitational Wave Signals?” In: (Dec. 2025). arXiv: 2512.17550 [gr-qc].

[180] R. Abbott et al. “GW190521: A Binary Black Hole Merger with a Total Mass of $150M_{\odot}$”. In: *Phys. Rev. Lett.* 125.10 (2020), p. 101102. DOI: 10.1103/PhysRevLett.125.101102. arXiv: 2009.01075 [gr-qc].

[181] R. Abbott et al. “Properties and Astrophysical Implications of the 150 $M_{\odot}$ Binary Black Hole Merger GW190521”. In: *Astrophys. J. Lett.* 900.1 (2020), p. L13. DOI: 10.3847/2041-8213/aba493. arXiv: 2009.01190 [astro-ph.HE].

[182] Isobel M. Romero-Shaw et al. “GW190521: orbital eccentricity and signatures of dynamical formation in a binary black hole merger signal”. In: *Astrophys. J. Lett.* 903.1 (2020), p. L5. DOI: 10.3847/2041-8213/abbe26. arXiv: 2009.04771 [astro-ph.HE].

[183] Rossella Gamba et al. “GW190521 as a dynamical capture of two nonspinning black holes”. In: *Nature Astron.* 7.1 (2023), pp. 11–17. DOI: 10.1038/s41550-022-01813-w. arXiv: 2106.05575 [gr-qc].

[184] Juan Calderón Bustillo et al. “GW190521 as a Merger of Proca Stars: A Potential New Vector Boson of 8.7×10^{-13} eV”. In: *Phys. Rev. Lett.* 126.8 (2021), p. 081101. DOI: 10.1103/PhysRevLett.126.081101. arXiv: 2009.05376 [gr-qc].

[185] V. De Luca et al. “GW190521 Mass Gap Event and the Primordial Black Hole Scenario”. In: *Phys. Rev. Lett.* 126.5 (2021), p. 051101. DOI: 10.1103/PhysRevLett.126.051101. arXiv: 2009.01728 [astro-ph.CO].

[186] B. P. Abbott et al. “Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA”. In: *Living Rev. Rel.* 21.1 (2018), p. 3. DOI: 10.1007/s41114-020-00026-9. arXiv: 1304.0670 [gr-qc].

[187] *Noise curves used for Simulations in the update of the Observing Scenarios Paper*. URL: https://dcc.ligo.org/LIGO-T2000012/public.

[188] Geraint Pratten et al. “Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes”. In: *Phys. Rev. D* 103.10 (2021), p. 104056. DOI: 10.1103/PhysRevD.103.104056. arXiv: 2004.06503 [gr-qc].

[189] Karl Wette. “SWIGLAL: Python and Octave interfaces to the LALSuite gravitational-wave data analysis libraries”. In: *SoftwareX* 12 (2020), p. 100634. DOI: 10.1016/j.softx.2020.100634.

[190] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2017. arXiv: 1412.6980 [cs.LG]. URL: https://arxiv.org/abs/1412.6980.

[191] Ilya Loshchilov and Frank Hutter. “SGDR: Stochastic Gradient Descent with Warm Restarts”. In: *arXiv e-prints*, arXiv:1608.03983 (Aug. 2016), arXiv:1608.03983. DOI: 10.48550/arXiv.1608.03983. arXiv: 1608.03983 [cs.LG].

[192] Vijay Varma et al. “Surrogate models for precessing binary black hole simulations with unequal masses”. In: *Phys. Rev. Research.* 1 (2019), p. 033015. DOI: 10.1103/PhysRevResearch.1.033015. arXiv: 1905.09300 [gr-qc].

[193] LIGO Scientific, Virgo, and KAGRA Collaboration. *GW231123: a Binary Black Hole Merger with Total Mass 190-265 $M_\odot$ — Data Release*. July 2025. DOI: 10.5281/zenodo.15832843. URL: https://doi.org/10.5281/zenodo.15832843.

[194] Skipper Seabold, Josef Perktold, et al. “Statsmodels: econometric and statistical modeling with python.” In: *SciPy* 7.1 (2010), pp. 92–96.

[195] `GWMAT`: Gravitational Waves Microlensing Analysis Tools. https://git.ligo.org/anuj.mishra/gwmat. 2023.

[196] Jose María Ezquiaga, Wayne Hu, and Rico K. L. Lo. “Identifying strongly lensed gravitational waves through their phase consistency”. In: *Phys. Rev. D* 108.10 (2023), p. 103520. DOI: 10.1103/PhysRevD.108.103520. arXiv: 2308.06616 [astro-ph.CO].

[197] Neil J. Cornish and Tyson B. Littenberg. “BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches”. In: *Class. Quant. Grav.* 32.13 (2015), p. 135012. DOI: 10.1088/0264-9381/32/13/135012. arXiv: 1410.3835 [gr-qc].

[198] Sophie Hourihane et al. “Accurate modeling and mitigation of overlapping signals and glitches in gravitational-wave data”. In: *Phys. Rev. D* 106.4 (2022), p. 042006. DOI: 10.1103/PhysRevD.106.042006. arXiv: 2205.13580 [gr-qc].

[199] D. Davis et al. “Improving the Sensitivity of Advanced LIGO Using Noise Subtraction”. In: *Class. Quant. Grav.* 36.5 (2019), p. 055011. DOI: 10.1088/1361-6382/ab01c5. arXiv: 1809.05348 [astro-ph.IM].

[200] D. Davis et al. “Subtracting glitches from gravitational-wave detector data during the third LIGO-Virgo observing run”. In: *Class. Quant. Grav.* 39.24 (2022), p. 245013. DOI: 10.1088/1361-6382/aca238. arXiv: 2207.03429 [astro-ph.IM].

[201] J. Veitch et al. “Robust parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library”. In: *Phys. Rev. D* 91 (4 2015), p. 042003. DOI: 10.1103/PhysRevD.91.042003. arXiv: 1409.7215 [gr-qc].

[202] *Noise Curves for use in simulations pre-O4*. URL: https://dcc.ligo.org/T2200043-v3/public.