

# Deep Learning Based Search and Parameter Estimation of Gravitationally Lensed Gravitational Waves

Saqlain Afroz (20MS230)

Indian Institute of Science Education and Research Kolkata

Department of Physical Sciences

November 2024

*Masters Thesis Report for 9th Semester*



---

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Gravitational Lensing</b>	<b>4</b>
2.1	Basics of Gravitational Lensing . . . . .	4
2.1.1	Point Mass Lens . . . . .	4
2.1.2	Time Delay . . . . .	6
2.1.3	Extended Mass Distribution . . . . .	6
2.2	Gravitational Lensing of Light . . . . .	7
2.2.1	Lens Equation . . . . .	7
2.2.2	Amplification . . . . .	8
2.2.3	Caustics and Critical Curves . . . . .	9
2.2.4	Fermat's Principle . . . . .	9
2.3	Differences in Lensing of EM waves vs Gravitational Waves . . . . .	10
2.4	Gravitational Lensing of Gravitational Waves . . . . .	10
2.4.1	Gravitational Lensing of GWs for Point Mass Lens . . . . .	11
2.5	Effect of lensing in Gravitational Waves . . . . .	13
<b>3</b>	<b>Methodology</b>	<b>13</b>
3.1	Parameter Estimation of Lensed Gravitational Waves . . . . .	13
3.1.1	Model Comparison: Microlensed vs Unlensed . . . . .	13
3.2	Machine Learning Approach for Parameter Estimation . . . . .	15
3.2.1	Finding Posterior Distribution: Introduction to NPEs . . . . .	15
3.2.2	Model architecture . . . . .	16
3.2.3	Preparation of Data . . . . .	17
3.2.4	Training Neural Network Model . . . . .	17
3.3	Bayesian Neural Network Approach . . . . .	19
3.4	Preparation of Data and Training . . . . .	20
<b>4</b>	<b>Results</b>	<b>20</b>
4.1	For Neural Posterior Estimation Model . . . . .	20
4.2	For Bayesian Neural Network Model . . . . .	21
<b>5</b>	<b>Conclusion and Future Prospects</b>	<b>21</b>

---

## Abstract

Gravitational waves (GWs), ripples in spacetime predicted by Einstein's theory of general relativity, offer a unique window into the cosmos. As detectors achieve greater sensitivity, the phenomenon of gravitational lensing, where massive objects bend these waves, is increasingly significant. This study addresses the critical need for accurate search and parameter estimation of gravitationally lensed GWs to avoid biases introduced by neglecting lensing effects. Focusing on microlensing, we first establish the impact of lensing on GW waveforms and parameter estimation using Bayesian analysis. Traditional methods, while accurate, are computationally intensive, motivating the use of machine learning approaches. We employ two methods, namely Neural Posterior Estimation (NPE) with conditional normalizing flows to approximate posterior distributions, enabling rapid and reliable parameter estimation, and Bayesian Neural Network (BNN). By training on simulated datasets of lensed waveforms, we demonstrate the efficacy of NPE method in recovering source properties, significantly reducing computational overhead without compromising accuracy. Results validate the approach for microlensing and highlight avenues for extending this framework to more complex lensing scenarios. This work paves the way for integrating advanced machine learning tools into gravitational wave astronomy, improving the detection and characterization of lensed signals in the upcoming era of high-precision GW observatories. BNN work is still under progress but we get good results which tell us that we can continue exploring the field. Additionally we have showed the power of Physics Informed Neural Network (PINN) to help solve complex physical problems by finding solution to the diffraction and integral equations.

## 1 Introduction

Gravitational Waves are ripple in the fabric of spacetime; they stretch and squeeze the spacetime fabric as they travel across the Universe. Gravitational Waves were predicted by Albert Einstein using his field equations,

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \kappa T_{\mu\nu} \quad (1)$$

by solving them under the condition of small perturbation in the background metric and considering the propagation in the empty space (thus implying  $T_{\mu\nu} = 0$ ). We get the final solution in the form of two polarizations of gravitational waves, namely the plus polarisation ( $h_+$ ) and the cross polarisation ( $h_\times$ ). The first direct detection of GWs were made using Laser Interferometer and Gravitational Wave Observatory (LIGO) and since then 90 events have been detected.

Gravitational Lensing (GL) is an another phenomena which have been predicted by Albert Einstein which explains why waves and light rays bend in the presence of a gravitational field [1]. With the development of detector being more sensitive to higher redshifts there is higher probability of detecting gravitationally lensed gravitational waves. So, we need to develop mechanism so that we can do proper inference on those GWs, because if we don't consider lensing effect, we will inherently introduce bias in our results. In this paper, we will first cover the basics of lensing, and how they differ in case of EM waves and Gravitational Waves; then we will introduce the concept of parameter estimation using Bayesian analysis and see what effects do lensing have in the waveform and their parameter estimation; then we will cover the machine learning approach we are using to do parameter estimation and lastly we present the results we have obtained so far, along with future prospects of our project.

## 2 Gravitational Lensing

Gravitational Waves and EM waves travel along null geodesic. And, GL is the theory of lightlike geodesic in Lorentzian manifold. Lensing is caused because the spacetime is curved due to a presence of massive object.

While working with Gravitational Lensing of Gravitational Waves, we need to consider certain aspects according to which we decide to either work in geometric optics or wave optics. We will introduce some basic concepts of lensing before going to description of geometrical and wave optics regime. While introducing the basic concepts we will use consider deflection of light rays which more or less is valid for gravitational waves.

### 2.1 Basics of Gravitational Lensing

We know that gravitational lensing is bending of light or GWs when there is a curvature in the spacetime. Before going to the much rigorous definitions we will first introduce some basic terms which will be helpful later on in our study, like *deflection angle*, which is the angle by which the light rays or GWs deflect when encountering a mass in the between the trajectory; *Shapiro time delay*, which is the time delay caused due to the presence of gravitational field; *Lens equation*, which governs the dynamics of lensing; *Amplification factor*, which tells that by what factor the light flux or GW amplitude is amplified due to lensing effect; *Fermat's Principle*, which tells that light travels along the path for which it takes least time, and then we will go on to discuss the specific case of lensing of GWs where we need to consider wave optics effects.

We will start with *Point mass lens* because it's easier to treat this particular case and we can easily extend this concept to more complicated lens system.

#### 2.1.1 Point Mass Lens

When light rays encounter a massive object they suffer some deflection which is measured as *deflection angle*. The deflection angle can be obtained from the knowledge of the trajectory of the light rays, which correspond to null geodesics and are hence solutions of the geodesic equations. We will consider Schwarzschild metric because we have spherical symmetry due to a point like mass and we also note that the trajectory of a light is defined by incident light direction and the mass location, thus we confine the light plane to be  $\theta = \pi/2$ . Hence, the line element that we get is as following:

$$ds^2 = \left(1 - \frac{2GM}{rc^2}\right) c^2 dt^2 - \left(1 - \frac{2GM}{rc^2}\right)^{-1} dr^2 - r^2 d\phi^2 \quad (2)$$

We can see that our metric has  $t$  and  $\phi$  as cyclic coordinates, which will give us two integrals of motion:

$$\frac{d}{d\lambda} \frac{\partial}{\partial \dot{\phi}} \left( \frac{ds}{d\lambda} \right)^2 = 0 \quad \longrightarrow \quad \implies r^2 \dot{\phi} = J \quad (3)$$

$$\frac{d}{d\lambda} \frac{\partial}{\partial \dot{t}} \left( \frac{ds}{d\lambda} \right)^2 = 0 \quad \longrightarrow \quad \implies \left(1 - \frac{2GM}{rc^2}\right) c^2 \dot{t} = A \quad (4)$$

where  $J$  and  $A$  are constants. As we are dealing with lightlike geodesics, we have

$$\left(1 - \frac{2GM}{rc^2}\right) c^2 \dot{t}^2 - \left(1 - \frac{2GM}{rc^2}\right)^{-1} \dot{r}^2 - r^2 \dot{\phi}^2 = 0 \quad (5)$$

We can take our affine parameter  $\lambda$  such that we get  $c^2 = A$  (without loss of generality). We substitute Eq. (3) and Eq. (4) in Eq. (5) to obtain:

$$\dot{r}^2 = 1 - \frac{J^2}{r^2} \left(1 - \frac{2GM}{rc^2}\right) \quad (6)$$

Now, we use Eq. (3) again to obtain the relation

$$d\phi = \frac{J}{r^2} \frac{dr}{\sqrt{1 - \frac{J^2}{r^2} \left(1 - \frac{2GM}{rc^2}\right)}} \quad (7)$$

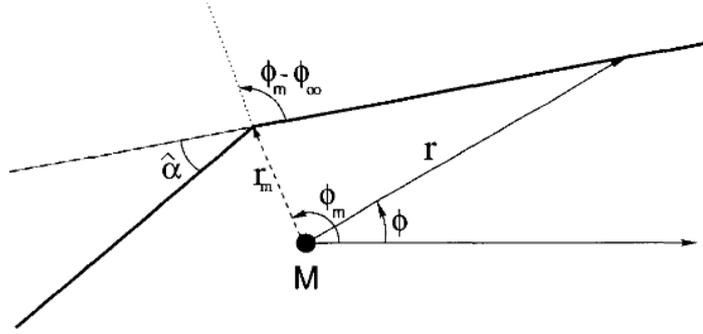


Figure 2: Deflection of light ray (Ref. [8])

At the time of closest approach to  $M$ ,  $\phi = \phi_m$ ,  $r = r_m$  and because there is not radial change at the time of deflection, thus we also have  $dr/d\phi = 0$ . Which gives us

$$J = \frac{r_m}{\sqrt{1 - \frac{2GM}{r_m c^2}}} \quad (8)$$

We do a substitution as  $x = r_m r$  and use the relation given in Eq. (8) to obtain

$$\phi_m - \phi_\infty = \int_0^1 \frac{dx}{\sqrt{1 - x^2 - \frac{2GM}{r_m c^2} (1 - x^3)}} \quad (9)$$

Where  $\phi_\infty$  is the asymptotic value of  $\phi$  very far in the past (the incident direction). When  $r_m \gg r_s = 2GM/c^2$ , upon solving the RHS we directly get

$$\phi_m - \phi_\infty = \frac{\pi}{2} + \frac{2GM}{r_m c^2} \quad (10)$$

The difference of the two asymptotic directions is twice of this quantity and differs from deflection angle by  $\frac{\pi}{2}$ . Thus we finally obtain

$$\hat{\alpha} = \frac{4GM}{r_m c^2} \quad (11)$$

### 2.1.2 Time Delay

Due to the curvature of spacetime we can have time delay in the arrival of light rays as compared to the case when there is no deflecting body. The time delay has two contributions, one known as geometrical time delay  $\delta t_{\text{geom}}$ , and the other one is gravitational time delay  $\delta t_{\text{grav}}$ , which is also known as *Shapiro time delay*, which arises from the fact that near higher gravitational field the time slows down and distance lengths are affected. To find Shapiro time delay we first substitute  $r = R \left[ 1 + \left( \frac{GM}{2Rc^2} \right)^2 \right]$  and then introduce Cartesian coordinate system to give us:

$$ds^2 = \left( \frac{1 - GM/2Rc^2}{1 + GM/2Rc^2} \right)^2 c^2 dt^2 - \left( 1 + \frac{GM}{2Rc^2} \right)^4 (dx^2 + dy^2 + dz^2) \quad (12)$$

We know that in weak field limit we have  $g_{00} = 1 + \frac{2\varphi}{c^2}$ . Thus, we get,

$$ds^2 = \left( 1 + \frac{2\varphi}{c^2} \right) c^2 dt^2 - \left( 1 - \frac{2\varphi}{c^2} \right) (dx^2 + dy^2 + dz^2) \quad (13)$$

For simplicity we consider light ray (for which  $ds = 0$ ) travelling along z-direction, thus we have

$$\begin{aligned} t_f - t_i &\simeq \frac{1}{c} \int_{z_i}^{z_f} \left( 1 - \frac{2\varphi}{c^2} \right) dz \\ \delta t_{\text{grav}} &= -\frac{2}{c^3} \int_{z_i}^{z_f} \varphi(z) dz \end{aligned} \quad (14)$$

### 2.1.3 Extended Mass Distribution

In more realistic scenario we have an extended mass distribution which acts as lens. Let's suppose that the density is given as  $\rho(\vec{x})$ . We make an assumption that the extension of this mass distribution is small in the direction of propagation of light as compared to perpendicular direction of light path, i.e. we work in the thin lens approximation. We project this mass density onto the plane orthogonal to light ray direction, to get

$$\Sigma(\vec{\xi}) = \int dz \rho(\vec{\xi}, z) \quad (15)$$

where  $\vec{\xi}$  is the two dimensional vector in the lens plane and  $z$  is the coordinate in the direction of propagation of light. As we are working in the weak field limit thus we can write the total deflection angle as:

$$\vec{\alpha}(\vec{\xi}) = \frac{4G}{c^2} \int d^2 \xi' \Sigma(\vec{\xi}') \frac{\vec{\xi} - \vec{\xi}'}{|\vec{\xi} - \vec{\xi}'|^2} \quad (16)$$

Suppose we have gravitational potential  $\varphi(\vec{\xi}, z)$  then we can write the projected gravitational potential (projected onto the lens plane) as,

$$\psi(\vec{\xi}) = \int dz \varphi(\vec{\xi}, z) \quad (17)$$

and using Poisson's equation we can write

$$\nabla_{\xi}^2 \psi(\vec{\xi}) = 4\pi G \Sigma(\vec{\xi}) \quad (18)$$

We know that the Green's function which satisfies 2D Laplacian is given by  $\mathcal{G} = \ln|\vec{\xi} - \vec{\xi}'|$  which gives the projected potential as,

$$\psi(\vec{\xi}) = 2G \int d^2\xi' \Sigma(\vec{\xi}') \ln|\vec{\xi} - \vec{\xi}'| \quad (19)$$

Therefore, the deflection angle can be written, by using Eq. (16) and Eq. (19), as

$$\vec{\hat{\alpha}}(\vec{\xi}) = \frac{2}{c^2} \nabla_{\xi} \psi(\vec{\xi}) \quad (20)$$

In case of extended object if we have spherical symmetry then, we can apply Gauss theorem, to get that the deflection angle is proportional to the mass enclosed in the impact parameter,  $\xi$ .

## 2.2 Gravitational Lensing of Light

### 2.2.1 Lens Equation

In Fig. 3, we can see that the source is located at an angular distance  $\beta$ , the image position is at  $\theta$ , the deflection suffered is  $\hat{\alpha}$ , the distance from source to lens is  $D_{LS}$ , from source to observer is  $D_{OS}$  and that from observer to lens is  $D_{OL}$ .

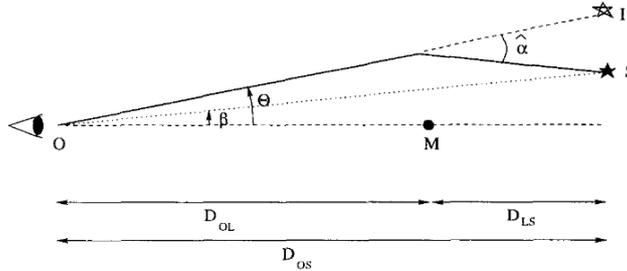


Figure 3: Geometry of point-like lens system [8]

From pure geometry we get the relation,

$$\beta = \theta - \alpha \quad (21)$$

where we define  $\alpha$  as the scaled deflection angle such that,  $\alpha = \hat{\alpha} \frac{D_{LS}}{D_{OS}}$ . Eq. (21) is known as the *lens equation*.

We can also write this lens equation after doing some substitution and defining some constants as,

$$\alpha = \frac{D_{LS}}{D_{OS} D_{OL}} \frac{4GM}{c^2 \theta} \quad (22)$$

and Einstein angle,

$$\theta_E = \sqrt{\frac{D_{LS}}{D_{OS}D_{OL}} \frac{4GM}{c^2}} \quad (23)$$

thus to write our lens equation as

$$\theta^2 - \beta\theta - \theta_E^2 = 0 \quad (24)$$

The solution of this lens equation gives us two values for  $\theta$ , which implies that in the case of point like lens we have two images. The significance of this Einstein's angle is that in the case if source lies exactly behind the lens then we will have a ring of image of the source with radial angle equal to Einstein's angle and also if the source is marginally displaced wrt. the optical axis as compared to the Einstein's angle then the separation between two images is of the order of Einstein's angle.

For an extended lens system we can write the lens equation in vectorial form as,

$$\vec{\beta} = \vec{\theta} - \vec{\alpha} \quad (25)$$

Using Eq. (20), we can rewrite the lens equation as,

$$\vec{\beta} = \vec{\theta} - \vec{\nabla}_{\theta} \Psi(\vec{\theta}) \quad (26)$$

where we have defined the scaled gravitational potential as,

$$\Psi = \frac{2}{c^2} \frac{D_{LS}}{D_{OS}D_{OL}} \psi \quad (27)$$

From Eq. (25) we can write the matrix of lens mapping as

$$\mathcal{T}_{ij} = \frac{\partial \beta_i}{\partial \theta_j} = \left( \delta_{ij} - \frac{\partial^2 \Psi}{\partial \theta_i \partial \theta_j} \right) \quad (28)$$

Recalling that Laplacian of  $\Psi$  is proportional to the surface density  $\Sigma(\vec{\theta})$ , we can write

$$\text{Tr} \frac{\partial^2 \Psi}{\partial \theta_i \partial \theta_j} = \nabla_{\theta}^2 \Psi = 2 \frac{\Sigma(\vec{\theta})}{\Sigma_{\text{cr}}} \quad (29)$$

where,  $\Sigma_{\text{cr}}$ , or the critical density is defined as,

$$\Sigma_{\text{cr}} \equiv \frac{c^2 D_{OS}}{4\pi G D_{OL} D_{LS}} \quad (30)$$

### 2.2.2 Amplification

The flux of EM waves that we receive is proportional to surface brightness and the solid angle it subtends. Now because the surface brightness is conserved and the solid angle it subtends changes due to lensing, we can have amplification or de-amplification. Let's suppose that the solid angle subtended when there is no lens is equal to  $d\Omega_0$  and the one subtended in the presence of lens is known denoted as  $d\Omega$ . Thus, the magnification is written as,

$$A = \frac{d\Omega}{d\Omega_0} \quad (31)$$

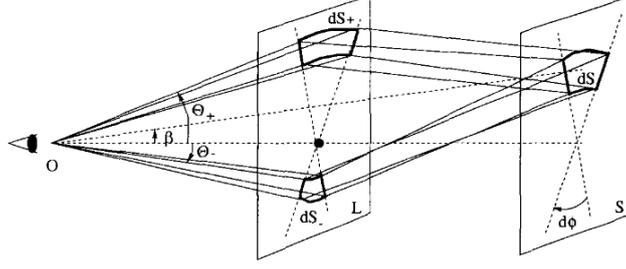


Figure 4: Ray path for lensed images [8]

From Fig. 4 we can write  $d\Omega_0 = \beta d\phi d\beta$  and for the two images we get  $d\Omega_{\pm} = \theta_{\pm} d\phi d\theta_{\pm}$ . Therefore we get the magnification as,

$$A_{\pm} = \frac{d\Omega_{\pm}}{d\Omega_0} = \frac{\theta_{\pm} d\theta_{\pm}}{\beta d\beta} \quad (32)$$

We can define Jacobian of the lens mapping as  $J = \det(\partial\vec{\beta}/\partial\vec{\theta})$ , and write  $A = J^{-1}$ .

### 2.2.3 Caustics and Critical Curves

Before we go on to further details of lensing, we will introduce the concepts of caustics and curves. From the relation of magnification with Jacobian, we can say that there might be some points for which the magnification diverges. The locus of these points, in the source plane, at which the amplification diverges is known as *Caustics*. The map of these caustics in the lens plane is known as critical lines.

In reality the amplification is not infinity because the overall amplification is weighted average of surface brightness. Caustics and critical lines are important because it tells us how many number of images will be formed wrt. the source position in the sky. If the source is far away from the lens, we will have only one image, but as it gets closer, we will have formation of multiple images. Every time the source crosses the caustics, we get extra pair of images [4]. Therefore, the number of images formed in lensing is always odd (for a non-singular lens), and this is known as Burke's theorem.

### 2.2.4 Fermat's Principle

We can write the lens equation given in Eq. (26) as,

$$\vec{\nabla}_{\theta} \left( \frac{1}{2}(\vec{\theta} - \vec{\beta})^2 - \Psi \right) = 0 \quad (33)$$

The first term in the parenthesis is the geometrical time delay while the second terms is gravitational time delay. If we denote the total time delay as  $\delta t$ , then we can write the lens equation as,

$$\vec{\nabla}_{\theta}(\delta t) = 0 \quad (34)$$

This is known as *Fermat's Principle*. The image is formed along the direction  $\vec{\theta}$  where the time delay has extremum.

### 2.3 Differences in Lensing of EM waves vs Gravitational Waves

Before going on to discuss about gravitational lensing of GWs, we first discuss about the differences in gravitational lensing of EM waves and GWs, which are given as following:

- Gravitational waves are coherent in nature so they can show signs of interference.
- We calculate the strain of gravitational waves as compared the flux which we do in case of EM waves. There the flux is changed by a magnitude of  $\mu$ , whereas the amplitude of GW is changed by a magnitude of  $\sqrt{\mu}$ , where  $\mu$  is the magnification caused due to lensing.
- Gravitational waves have negligible interaction with the intervening matter. Hence, lens objects even in optically thick region can be detected by lensing.
- The current ground based detectors function in the frequency range of  $\sim 10^1 - 10^4$  Hz whereas EM observations are made in the frequency range of  $\sim 10^6 - 10^{20}$  Hz. Which implies that the wavelength of GWs are much longer than EM waves and thus GWs can have diffraction effect arising when they encounter an object.

### 2.4 Gravitational Lensing of Gravitational Waves

We consider the propagation of gravitational waves in the presence of the gravitational potential,  $U(U(r) \ll 1)$ , of a lens object. The background metric in this case is given as [2]

$$ds^2 = -(1 + 2U)dt^2 + (1 - 2U)dr^2 = g_{\mu\nu}^{(B)} dx^\mu dx^\nu \quad (35)$$

We now introduce a perturbation  $h_{\mu\nu}$  in the background metric, so the total metric which we get is,

$$g_{\mu\nu} = g_{\mu\nu}^{(B)} + h_{\mu\nu} \quad (36)$$

Under the transverse traceless Lorentz gauge condition of  $h_{\nu\mu;\nu} = 0$  and  $h_{\mu}^{\mu} = 0$  we have

$$h_{\mu\nu;\alpha}^{\alpha} + 2R_{\alpha\mu\beta\nu}^{(B)} h^{\alpha\beta} = 0 \quad (37)$$

and if the wavelength is smaller than the curvature of the spacetime, we have  $h_{\mu\nu;\alpha}^{\alpha} = 0$ , thus using Eikonal approximation we get [6]

$$h_{\mu\nu} = \phi e_{\mu\nu} \quad (38)$$

where  $e_{\mu\nu}$  is the polarization tensor which is transported parallelly along the null geodesic. Then the change of the polarization tensor by gravitational lensing is of the order of  $U(\ll 1)$  which is very small in our observational situation, and hence we can regard the polarization tensor as a constant. Thus, we treat the scalar wave  $\phi$ , instead of the gravitational wave  $h_{\mu\nu}$ , propagating through the curved space-time. The propagation equation of the scalar wave in frequency domain is,

$$(\nabla^2 + \omega^2)\tilde{\phi} = 4\omega^2 U\tilde{\phi} \quad (39)$$

We define Amplification factor as following,

$$F(f) = \frac{\tilde{\phi}^L(f)}{\tilde{\phi}(f)} \quad (40)$$

We can solve Eq. (39) using Kirchhoff Integral as given in [4], to obtain,

$$F(f) = \frac{D_S \xi_0^2}{D_L D_{LS}} \frac{f}{i} \int d^2 \mathbf{x} \exp [2\pi i f t_d(\mathbf{x}, \mathbf{y})] \quad (41)$$

where  $\xi_0$  is an arbitrary length scale,  $\mathbf{x} = \xi/\xi_0$ ,  $\mathbf{y} = \eta D_L/\xi_0 D_S$ , also  $\xi = \theta D_{OL}$  and  $\eta = \beta D_{OS}$  (note that  $D_{OL}$  and  $D_{OS}$  have the same meaning as given in Section 2.2.1), and  $t_d$  is the time delay of GW, given as,

$$t_d(\mathbf{x}, \mathbf{y}) = \frac{D_{OS} \xi_0^2}{D_{OL} D_{LS}} \left[ \frac{1}{2} |\mathbf{x} - \mathbf{y}|^2 - \psi(\mathbf{x}) + \phi_m(y) \right] \quad (42)$$

When  $ft_d \gg 1$  (also known as geometrical optics limit, or strong lensing) the integrand in the Eq. (41) is highly oscillatory and only the stationary points of  $t_d$  (or  $\delta t$  from Eq. (34)) contribute to the integral in Eq. (41). Therefore, the integral on the lens plane (in Eq. (41)) is reduced to the sum over these images, [7] as

$$F(f) = \sum_j |\mu_j|^{1/2} \exp [2\pi i f t_d - i\pi n_j] \quad (43)$$

where  $\mu_j$  is the magnification of the  $j$ -th image and it is given as  $\mu_j = 1/\det(\partial\beta/\partial\theta)$ , and also  $n_j$  is the Morse index where  $n_j = 0, 1/2$ , and 1 represents minima, saddle point and maxima, respectively.

#### 2.4.1 Gravitational Lensing of GWs for Point Mass Lens

For point mass lens the surface mass density is expressed as  $\Sigma(\xi) = M_L \delta^2(\xi)$  where  $M_L$  is the lens mass, and according to Eq. (29) the non-dimensional deflection potential is  $\psi(x) = \ln x$ . According to [3], we can do analytical integration of 41, to obtain,

$$F(f) = \exp \left[ \frac{\pi\omega}{4} + i\frac{\omega}{2} \left( \ln \frac{\omega}{2} - 2\phi_m(y) \right) \right] \times \Gamma \left( 1 - \frac{i}{2}\omega \right) {}_1F_1 \left( \frac{i}{2}\omega, 1; \frac{i}{2}\omega y^2 \right) \quad (44)$$

where  ${}_1F_1$  is the confluent hypergeometric function,  $\omega = 8\pi M_{Lz} f$ ,  $\phi_m(y) = (x_m - y)^2/2 - \ln x_m$ , with  $x_m = [y + (y^2 + 4)^{1/2}]/2$  and  $M_{Lz}$  is the redshifted lens mass.

In case of  $ft_d \sim 1$  the interference effects become dominant and we need to consider wave optics. This leads to frequency dependent modulations of the signal, which case of point mass lens is governed by Eq. (44). This is also known as *microlensing of GWs*.

The GWs under the effect of microlensing creates a beating pattern which we can see in the Fig. 5(a) in Time Domain and in Fig. 5(b) in Frequency domain.

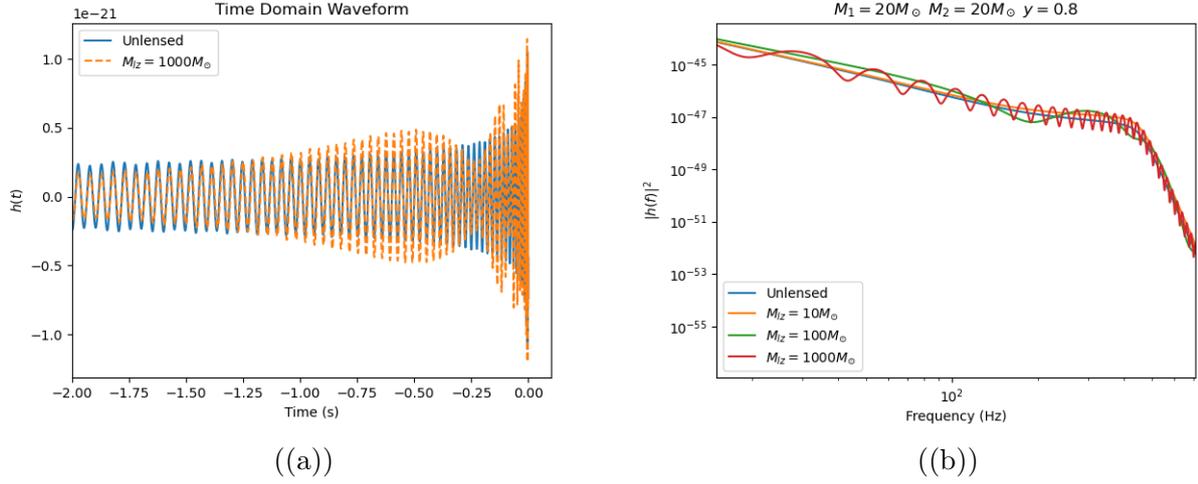


Figure 5: Microlensing of GWs

However, in case of  $ft_d \gg 1$ , Eq. (43) can be approximated for point mass lens to

$$F(f) = |\mu_+|^{1/2} - i|\mu_-|^{1/2}e^{i2\pi ft_d} \quad (45)$$

where  $|\mu_{\pm}| = 1/2 \pm (y^2 + 2)/(2y\sqrt{y^2 + 4})$  is the magnification of each image and time delay is as given in Eq. (42). This is also known as *Strong Lensing of GWs*.

For an illustrative purpose we can see the gravitational wave under strong lensing (for certain configuration of lens mass and impact parameter) as given in Fig. 6, that the first signal is magnified while the second signal is demagnified and they both arrive with a certain time delay.

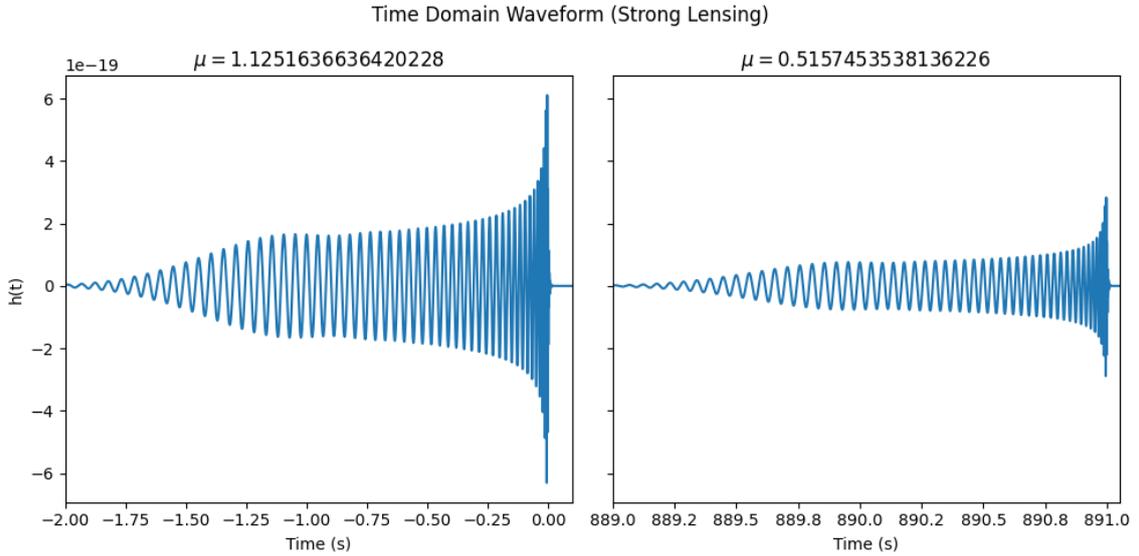


Figure 6: Strong Lensing of Gravitational Waves

## 2.5 Effect of lensing in Gravitational Waves

When we have strong lensing of gravitational waves, the amplitudes get amplified. Now the amplitude of gravitational wave depends on mass of the binary and luminosity distance. If the masses of the binary are high then we will have higher amplitude and vice-versa. Similarly, if the luminosity distance is small than also the amplitude is higher. So we can see that in case of strong lensing these two parameters get mostly affected, if we don't be careful while doing parameter estimation. However, for microlensing the parameters that get affected are not only masses and luminosity distance, but the spins and other parameters get affected as well [11]. We will see much about this in Section 3.1.1.

# 3 Methodology

## 3.1 Parameter Estimation of Lensed Gravitational Waves

Given a detection, Bayesian inference is used to characterize the originating source. This is based on having models for the signals and the detector noise. For gravitational waves, signal models take the form of waveform predictions  $h(\theta)$  depending on the source parameters  $\theta$  (masses, location, etc.). Waveform models are based on solutions to Einstein's equations (and any relevant matter equations) for the two-body dynamics and gravitational radiation, using a combination of numerical-relativity and perturbative calculations and phenomenological fitting. Detector noise is typically modelled as stationary and Gaussian, with some spectrum which can be estimated empirically. Together, these "forward" models give rise to the likelihood  $\mathcal{L}(d|\theta)$  for the observed strain data  $d$ , which is assumed to consist of a signal plus noise. With the choice of a prior  $\pi(\theta)$  over parameters, the posterior distribution is given via Bayes' theorem,

$$p(\theta|d) = \frac{\mathcal{L}(d|\theta)\pi(\theta)}{\mathcal{Z}} \quad (46)$$

The task of inference is to characterize the posterior by drawing samples from it. This can be accomplished with stochastic algorithms like Markov chain Monte Carlo (MCMC). The LVC have developed software tools such as LALInference and Bilby to carry this out. However, these algorithms are computationally expensive as they require many likelihood evaluations for each independent posterior sample  $\theta \sim p(\theta|d)$ , and each likelihood requires a waveform simulation. An analysis producing  $\sim 10^4$  independent samples typically requires millions of waveform evaluations and a total inference time of hours to months, depending on the signal duration and waveform model.

In this study we are using a large number of simulated data sets (with associated parameters), and use these to train a type of neural network known as a *normalizing flow* to approximate the posterior. The trained network can then generate new posterior samples extremely quickly once a detection is made. This bypasses the need to generate waveforms at inference time, thereby amortizing the expensive training costs over all future detections. The general approach of building such "surrogate" inverse models is called *neural posterior estimation*.

### 3.1.1 Model Comparison: Microlensed vs Unlensed

Before going on to do parameter estimation of lensed gravitational waves, we will first do testing of lensed and unlensed hypothesis. We generated lensed and unlensed waveform and found the Fitting factor at each point of  $(M_{lz}, y)$ , where the distribution in  $M_{lz}$  has been assumed to be LogUniform and PowerLaw (with  $\alpha = 0$ ) for the distribution of  $y$ .

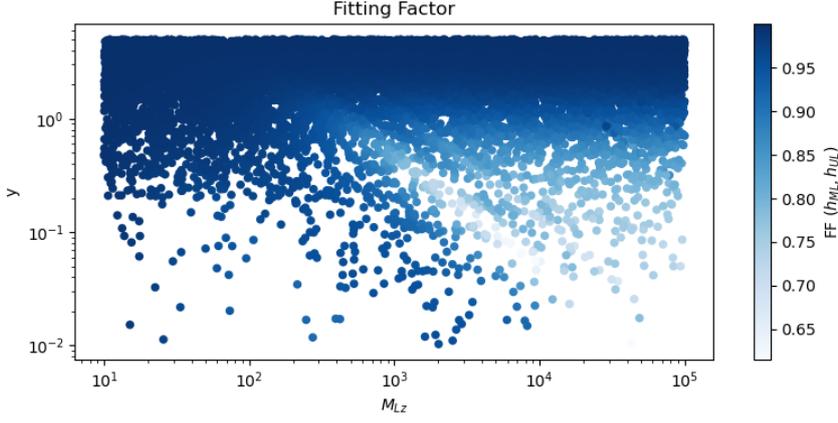


Figure 7: Fitting Factor

We can see in the plot that the fitting factor reduces as we go to higher lens mass and small impact parameter, implying that we need to be careful when trying to infer source parameter in this region. If we do not stay careful then we will introduce bias in our results of parameter estimation.

For this study a lensed waveform was generated using *bilby* and we tried to recover the source properties using two different models: an unlensed model and a lensed model. We are using *gwmat* [10], a Python library, to generate microlensed waveform. The parameter estimation done for each case is given below:

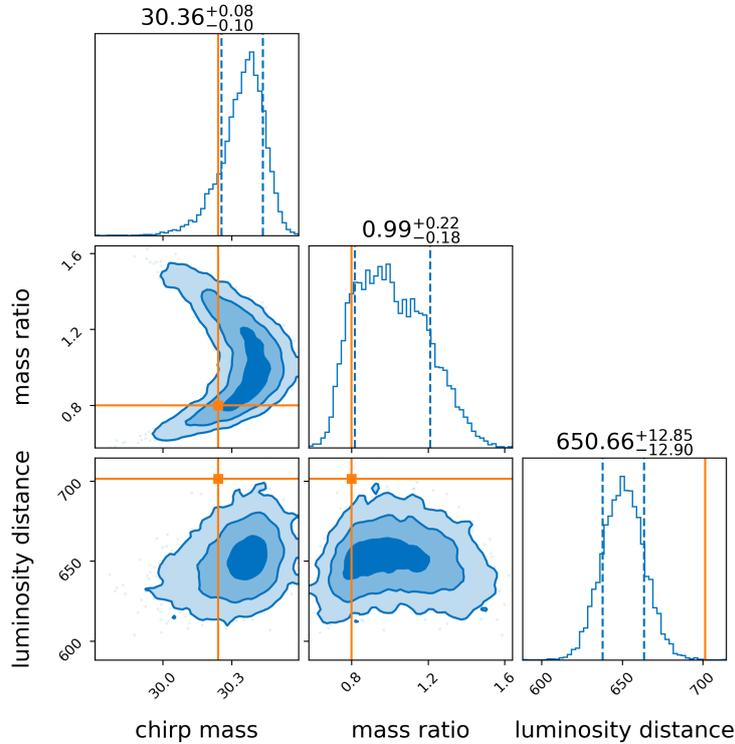


Figure 8: Parameter Estimation using Unlensed model

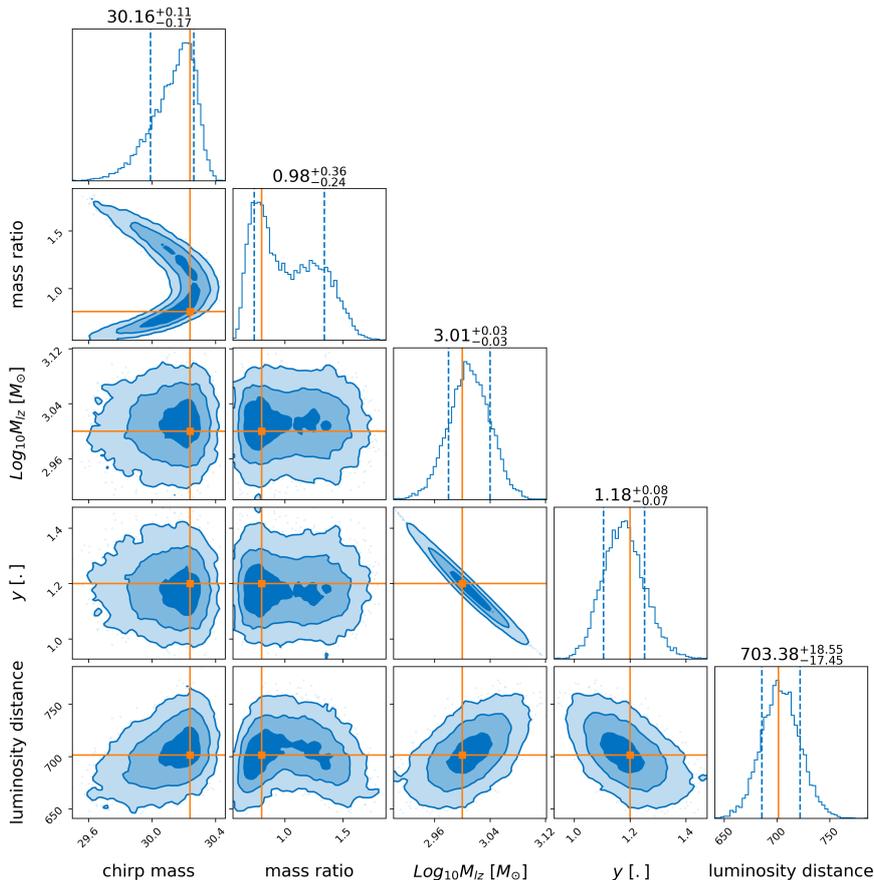


Figure 9: Parameter Estimation using Lensed model

The plots we got are in accordance with the statement we have written in Section 2.5, i.e. when we tried to recover the lensed waveform using the unlensed model, we can see that the chirp mass and mass ratio have been overestimated, while the luminosity distance has been underestimated, in Fig. 8. Whereas, in Fig. 9 we can see that all the parameters have been properly recovered when we tried to recover the lensed waveform with the lensed model.

This study gives the motivation of why do we need to consider lensing effects while doing parameter estimation.

## 3.2 Machine Learning Approach for Parameter Estimation

The traditional parameter estimation is computationally expensive and we need to employ other methods which can reduce the computational time but at the same time maintain the accuracy. For this purpose we will be using machine learning approach which has proven to be very helpful in the field of astrophysics recently.

### 3.2.1 Finding Posterior Distribution: Introduction to NPEs

For finding posterior distribution we will be using *Neural Posterior Estimation (NPE)*. The package we are using is known as *Dingo* [9]. The basic idea is to produce a large number of simulated data sets (with associated parameters), and use these to train a type of neural network known as *Normalizing Flows* to approximate the posterior. In normalizing flow, we start with a simple distribution, such as a uniform distribution, and transform it using a composition of bijective functions. This allows us to represent the posterior distribution. This process can be modelled using the formula given in Eq. (47).

$$\begin{aligned}
q_L(\theta_L)d\theta_L &= q_0(\theta_0)d\theta_0 \\
q_L(\theta_L) &= q_0(f^{-1}(\theta_L)) \prod_{i=1}^L \frac{\partial f^{-1}(\theta_L)}{\partial \theta_L}
\end{aligned} \tag{47}$$

We can also see how the Normalizing Flow works in an illustration in Fig. 10.

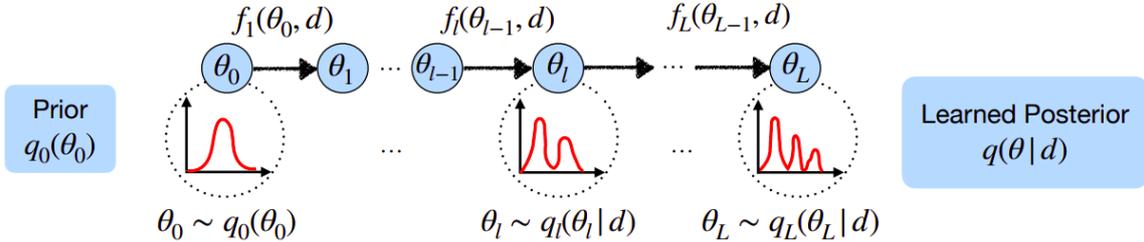


Figure 10: Normalizing Flows [13]

Here, each  $f_i$ 's are known as flows.

The trained network can then generate new posterior samples extremely quickly once a detection is made. This bypasses the need to generate waveforms at inference time, thereby amortizing the expensive training costs over all future detections. The general approach of building such “surrogate” inverse models is called NPE.

### 3.2.2 Model architecture

The architecture consists of two components, the embedding network which compresses the high-dimensional data to a lower dimensional feature vector, and the conditional normalizing flow which estimates the Bayesian posterior based on this feature vector.

The embedding network compresses the high-dimensional conditioning information (consisting of frequency domain strain and PSD data). The first layer of this network is initialized with an SVD matrix from a reduced basis built with non-noisy waveforms. This projection filters out the noise that is orthogonal to the signal manifold, and significantly simplifies the task for the neural network.

The initial compression layer is followed by a sequence of residual blocks consisting of dense layers for further compression.

We use the neural spline flow as a density estimator. This takes the output of the embedding network as context information and estimates the Bayesian posterior distribution.

The posterior distribution is represented in terms of an invertible normalizing flow, taking normally-distributed random variables  $u$  into posterior samples  $\theta$ . The flow therefore depends on a (compressed) representation of the noise properties  $S_n$  and the data  $d$ . The model architecture diagram is as shown in Fig. 11.

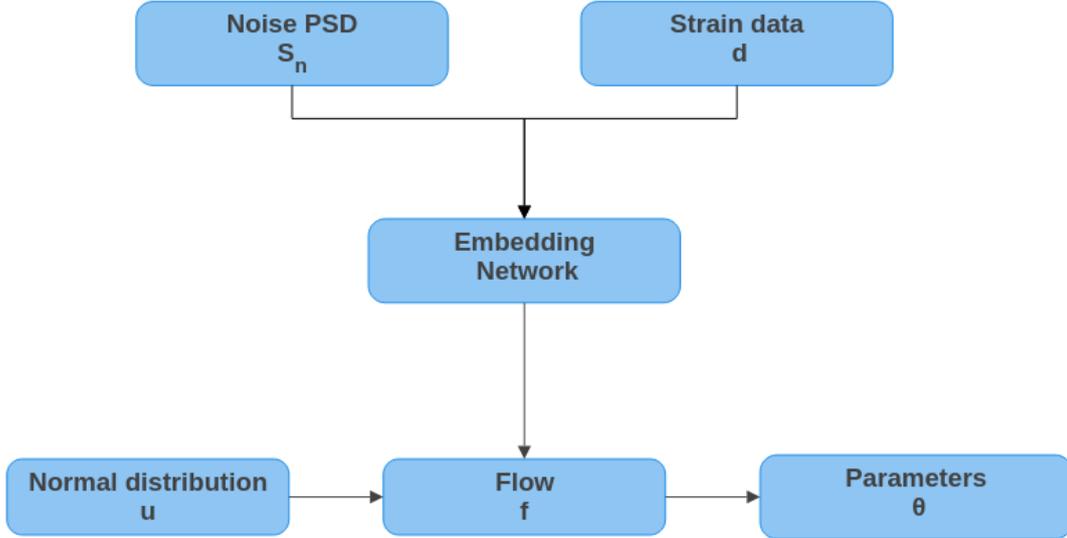


Figure 11: Neural Network Architecture

### 3.2.3 Preparation of Data

*Dingo* [9] provides the facility to generate a dataset of waveforms and store it in the form of a hdf5 file. We made modifications to the code so that it generates lensed gravitational waves (microlensed). The prior settings we used to generate the lensed waveforms is given in Table 1,

Parameter	Range	Distribution
$m_1, m_2$	$(10, 80) M_\odot$	Constraint
$\mathcal{M}$	$(15, 100)$	UniformInComponentChirpMass
$\mu$	$(0.125, 1)$	UniformInComponentMassRatio
$d_l$	$(100, 6000)$	Uniform
$\theta_{jn}$	$(0, \pi)$	Sine
$\psi$	$(0, 2\pi)$	Uniform
$\chi_1$	$(0, 0.9)$	AlignedSpin
$\chi_2$	$(0, 0.9)$	AlignedSpin
$\log M_{lz}$	$(-1, 5)$	LogUniform
$yl$	$(0.001, 5)$	PowerLaw ( $\alpha = 1$ )

Table 1: Prior distribution of the waveform dataset

Where the last two parameters correspond to the point lens system, the log of lens mass and impact parameter of the gravitational wave source with respect to lens. For this study we generated 2000 samples and conditioned it on  $O1$  PSD.

### 3.2.4 Training Neural Network Model

Before starting the training process, we need to define certain hyper-parameters, which are adjustable parameters in a neural network model that we set manually. They can be modified to optimize the model's performance. Some of the key hyper-parameters that we defined are as given in Table 2.

Hyper-parameter	Value
Number of flows	10
Activation Function	Elu
Learning Rate	0.0001
Learning Rate Scheduler	Cosine
Epochs	50

Table 2: Hyper-parameters for Neural Network

The conditional neural density estimator  $q(\theta|d)$  is initialized randomly and must be trained to become a good approximation to the posterior  $p(\theta|d)$ . To achieve this, one must specify a target loss function to minimize. In this project we are trying to minimize the Kullback-Leibler (KL) divergence of  $p$  from  $q$ ,

$$D_{KL}(p||q) = \int d\theta p(\theta|d) \log \frac{p(\theta|d)}{q(\theta|d)} \quad (48)$$

We use “forward KL divergence” and take expectation over data samples  $d \sim p(d)$  and we drop the numerator form log term as it does not depend on network parameters, to arrive at loss function which is given as:

$$\mathcal{L} = \int dd p(d) \int d\theta p(\theta|d) [-\log q(\theta|d)] \quad (49)$$

$$= \int d\theta p(\theta) \int dd p(d|\theta) [-\log q(\theta|d)] \quad (50)$$

We arrive at Eq.50 from Eq.49 using Bayes’ theorem  $p(d)p(\theta|d) = p(\theta)p(d|\theta)$

We can approximate the loss function as following:

$$\mathcal{L} \approx -\frac{1}{N} \sum_{i=1}^N \log q(\theta^{(i)}|d^{(i)}) \quad (51)$$

We then take gradient of  $\mathcal{L}$  with respect to our neural network parameters and minimize the Loss function by adjusting these parameters using an Optimizer known as “Adam Optimizer”, and learning rate as given in the equation below

$$\lambda_j := \lambda_j - \alpha \frac{\partial}{\partial \lambda_j} \mathcal{L} \quad (52)$$

where  $\lambda_j$  is the  $j$ -th parameter of the neural network and  $\alpha$  is the learning rate. We used a learning rate scheduler to ensure the training process converges after a certain number of epochs. In simple terms, if the initial learning rate is too high, the model may oscillate around the minimum of the loss function without settling. Conversely, if the learning rate is too low, convergence will be very slow, taking much longer to reach the minimum.

Training was done on Sarathi cluster of IUCAA and due to the small size of this model it took only  $\sim 25$  minutes, whereas full production scale model will take around 10 days with a single

A100 NVIDIA GPU as mentioned in [9], and without GPU it will take a month time to train. But when trained the inference is very fast. The train and test loss we obtained is given in Fig. 12

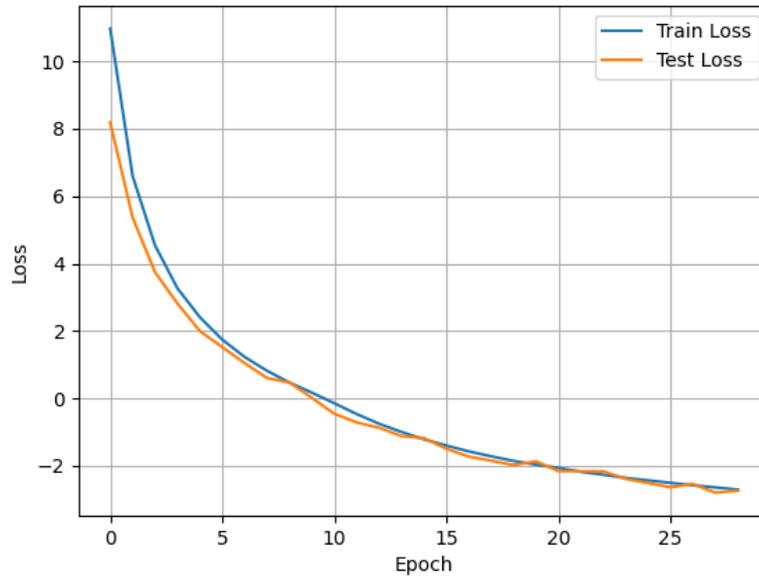


Figure 12: Train and Test Loss

The training and testing process show no sign of overfitting, in which case the train loss curve would have reduced while the test loss curve would have diverged, and also it shows no sign of underfitting, in which case neither the training loss nor the test loss would have reduced with respect to epochs.

### 3.3 Bayesian Neural Network Approach

Bayesian Neural Networks (BNNs) are a class of probabilistic models that integrate Bayesian inference into the traditional framework of neural networks. This integration allows BNNs to quantify uncertainty in model predictions, making them particularly useful in situations where uncertainty quantification is crucial, such as in medical diagnosis, autonomous systems, and financial forecasting. Traditional neural networks provide point estimates for predictions; however, they do not express uncertainty about the predictions, which can be a significant limitation, especially in high-risk or highly variable domains. Bayesian methods, on the other hand, offer a natural way to represent uncertainty by treating the weights of the network as probability distributions rather than fixed values.

In a standard neural network, the learning process involves adjusting the weights to minimize a loss function, typically using methods such as stochastic gradient descent (SGD). These models, however, assume that the learned weights are deterministic, which means they do not account for the fact that the optimal weights could vary due to noise in the data, model misspecification, or other sources of uncertainty. Bayesian Neural Networks address this limitation by introducing a probabilistic framework where the network's parameters are not fixed but are treated as random variables with associated probability distributions. This probabilistic treatment allows BNNs to not only make predictions but also provide a measure of uncertainty associated with those predictions.

The core idea of BNNs is rooted in Bayes' theorem, which allows for the updating of the distribution over the network's parameters based on observed data. The posterior distribution

over the parameters is typically intractable to compute exactly, leading to the need for approximate inference methods. Various techniques, such as Markov Chain Monte Carlo (MCMC), Variational Inference (VI), and Monte Carlo Dropout, have been proposed to approximate the posterior distribution. These methods allow the model to capture complex relationships between inputs and outputs while also providing uncertainty estimates that are valuable in many real-world applications.

One of the key benefits of Bayesian Neural Networks is their ability to handle small data or noisy data scenarios more effectively than traditional deep learning models. Because they can quantify uncertainty, BNNs can avoid overfitting by incorporating prior knowledge about the problem domain or adjusting their confidence in the predictions. This characteristic is particularly beneficial in domains with limited labeled data, where traditional models might fail to generalize well.

### 3.4 Preparation of Data and Training

We have generated 10 training dataset, which is small but we wanted to test the performance of our model. The waveform dataset is simulated using PyCBC [12] and the parameters which are sampled are the component masses and luminosity distance (note that we are not taking lensed data because this is just a test version). Then we are using 3 Bayesian Linear Layer using torchbnn<sup>1</sup>, with ReLU as 2 non-linear activation function. We are using Adam optimizer to adjust the neural network parameters so that our BNN model can get trained. The training epoch is set to 500.

## 4 Results

### 4.1 For Neural Posterior Estimation Model

The trained machine learning model was used to make inference on an injected data (as we have no evidence of lensed gravitational wave), which took around 5 seconds, and the results for the same are given below in Fig. 13. The inference was made on chirp mass, ratio, luminosity distance, impact parameter, and Log of redshifted lens mass.

As shown in Fig. 13, the posterior distributions are reasonably accurate, especially given that the neural network model used was relatively small. However, there is room for improvement, and further tuning of the hyper-parameters could enhance the model's performance, which will be done in nearby future.

---

<sup>1</sup>Link for BNN doc.

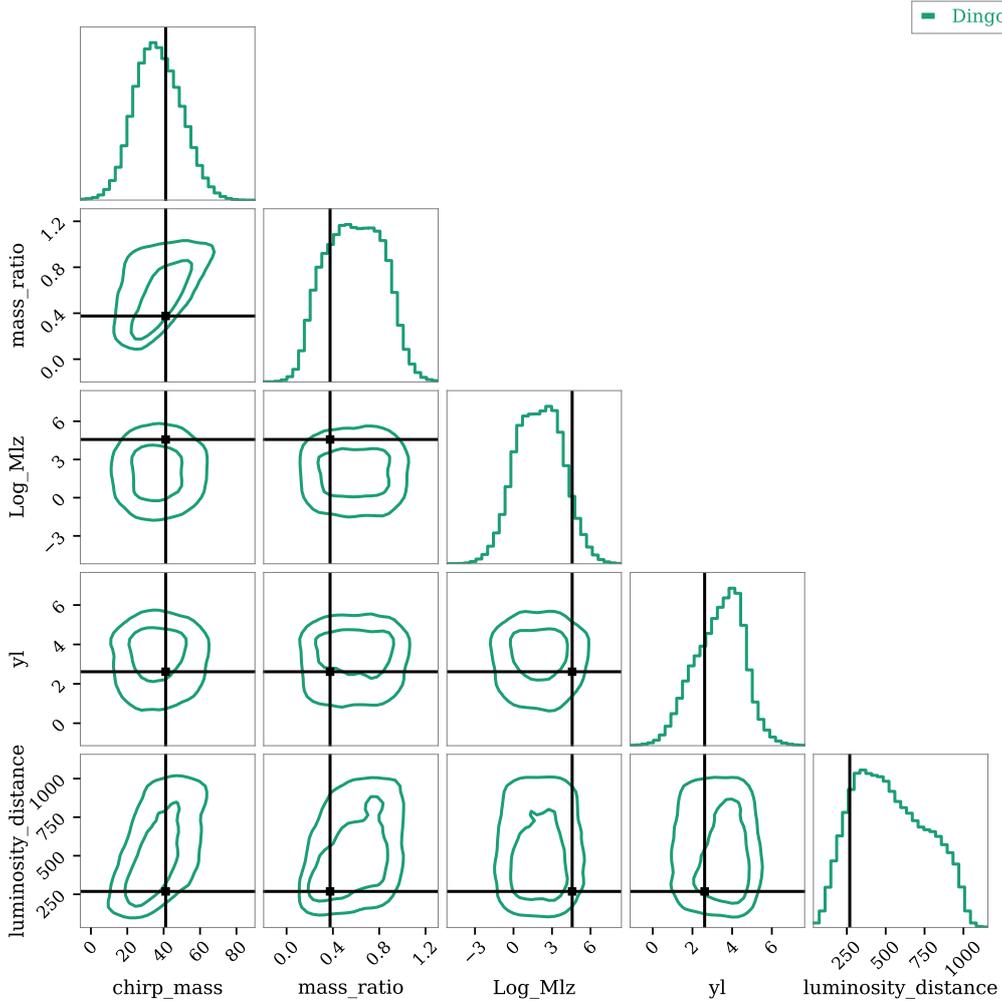


Figure 13: Corner Plot using Dingo (modified)

## 4.2 For Bayesian Neural Network Model

We created a simulated data with fixed parameters as: mass 1 = 10, mass 2 = 40, luminosity distance = 870, and then we made inference on that using our trained BNN model. The predicted values are  $20.18 \pm 2.50$ ,  $58.11 \pm 2.63$  and  $799.95 \pm 14.32$  for mass 1, mass 2 and luminosity distance, respectively. The results are close but not very accurate, and the possible reason for that could be less number of training data, or the fact that our BNN model architecture is simple. This is still a work in progress and we are working on it to improve the model.

## 5 Conclusion and Future Prospects

In this project, we introduced the concept of gravitational lensing and explained why it is important to consider lensing effects when conducting searches and parameter estimation for gravitational waves. We demonstrated this by comparing parameter estimation performed on lensed waveforms using both lensed and unlensed models. Our results highlighted how neglecting lensing can lead to biased outcomes if not handled carefully.

We then discussed the motivation for using machine learning in gravitational wave parameter estimation. While traditional methods like LALInference and Bilby are effective, machine learning offers the advantage of significantly reducing computation time while maintaining accuracy. To illustrate this, we presented the results obtained so far. To achieve these results,

we modified the Dingo code to generate lensed waveforms and trained a machine learning model on this data.

In this study, we focused on the point-mass lens model. However, in more realistic scenarios, we encounter extended objects like star clusters, galaxies, or galaxy clusters as lenses. Our future goal is to develop a model capable of performing parameter estimation for gravitational wave sources alongside more complex lensing systems, extending beyond the point-mass lens framework.

Also we plan to build a machine learning model that can help us solve Eq. (41), as it is impossible to solve it analytically and numerical solution for it, as provided by [5] and many other, is computationally expensive. We generated a plot to visualize the time complexity for computing amplification factor, using numerical approach, and it is given in Fig. 14.

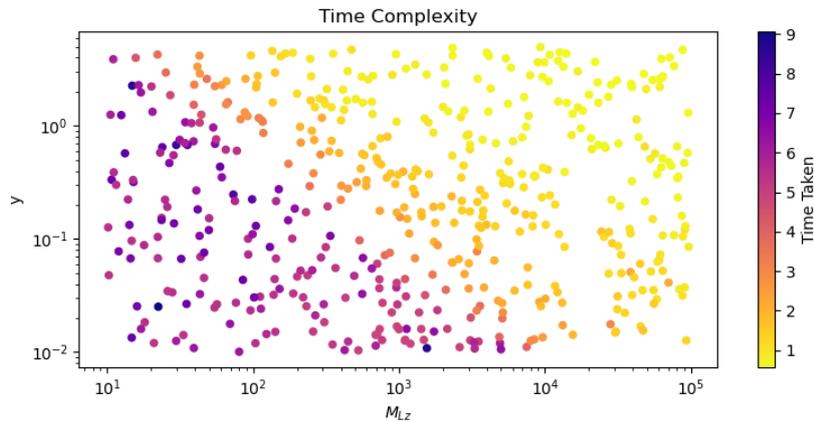


Figure 14: Computational Time Complexity for calculating Amplification Factor

We have built a Physics Informed Neural Network Model that can solve this problem for a case for which we do have analytical solution available (point mass lens) such that 71% of predicted amplification factor lie within an error margin of 15%. But we would like to improve this and build upon this to predict amplification factor for complex lens systems.

## Acknowledgment

I would like to thank my supervisor, Dr. Apratim Ganguly, for giving me the opportunity to work on this project. The experience was a significant learning curve, mainly in the context of gravitational lensing, and machine learning. I would like to thank Professor Rajesh Kumble Nayak for introducing me to the concepts of gravitational waves and their data analysis, which have profoundly shaped my understanding and research interests. I am extremely thankful to all my teachers at IISER Kolkata, who have shaped my academic and personal growth in many ways. Their guidance has taught me the essence of conducting research and instilled in me a deep passion for learning. Special thanks to my parents, who have always given me unflinching support and encouragement throughout this entire period, which has constantly built me up in strength and strength. I also would thank my friends for being on my side and keeping support going on and the precious memory of fun and camaraderie brightening up my days.

Also we acknowledge the use of IUCAA LDG cluster Sarathi for the computational/ numerical work.

## References

- [1] Albert Einstein. “Lens-Like Action of a Star by the Deviation of Light in the Gravitational Field”. In: *Science* 84.2188 (1936), pp. 506–507. DOI: 10.1126/science.84.2188.506. eprint: <https://www.science.org/doi/pdf/10.1126/science.84.2188.506>. URL: <https://www.science.org/doi/abs/10.1126/science.84.2188.506>.
- [2] Charles W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. San Francisco: W. H. Freeman, 1973. ISBN: 978-0-7167-0344-0, 978-0-691-17779-3.
- [3] P. C. Peters. “Index of refraction for scalar, electromagnetic, and gravitational waves in weak gravitational fields”. In: *Phys. Rev. D* 9 (8 Apr. 1974), pp. 2207–2218. DOI: 10.1103/PhysRevD.9.2207. URL: <https://link.aps.org/doi/10.1103/PhysRevD.9.2207>.
- [4] Peter Schneider, Jürgen Ehlers, and Emilio E. Falco. *Gravitational Lenses*. 1992. DOI: 10.1007/978-3-662-03758-4.
- [5] Andrew Ulmer and Jeremy Goodman. “Femtolensing: Beyond the semiclassical approximation”. In: *The Astrophysical Journal* 442 (Mar. 1995), p. 67. ISSN: 1538-4357. DOI: 10.1086/175422. URL: <http://dx.doi.org/10.1086/175422>.
- [6] Christian Baraldo, Akio Hosoya, and Takahiro T. Nakamura. “Gravitationally induced interference of gravitational waves by a rotating massive object”. In: *Phys. Rev. D* 59 (8 Mar. 1999), p. 083001. DOI: 10.1103/PhysRevD.59.083001. URL: <https://link.aps.org/doi/10.1103/PhysRevD.59.083001>.
- [7] T. T. Nakamura and S. Deguchi. “Wave Optics in Gravitational Lensing”. In: *Progress of Theoretical Physics* 133 (Jan. 1999), pp. 137–153. DOI: 10.1143/PTPS.133.137.
- [8] Valerio Bozza. “Silvia Mollerach, Esteban Roulet: Gravitational Lensing and Microlensing”. In: *General Relativity and Gravitation* 37 (July 2005), pp. 1335–1336. DOI: 10.1007/s10714-005-0117-9.
- [9] 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].
- [10] Anuj Mishra. *GWMat: A toolkit for gravitational wave data analysis*. GitLab repository. 2024. URL: <https://git.ligo.org/anuj.mishra/gwmat> (visited on 12/04/2024).
- [11] 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].
- [12] Alex Nitz et al. *gwastro/pycbc: v2.3.3 release of PyCBC*. Version v2.3.3. Jan. 2024. DOI: 10.5281/zenodo.10473621. URL: <https://doi.org/10.5281/zenodo.10473621>.
- [13] [https://indico.ph.tum.de/event/7722/contributions/9480/attachments/6368/8770/presentation\\_2024\\_09\\_23\\_print.pdf](https://indico.ph.tum.de/event/7722/contributions/9480/attachments/6368/8770/presentation_2024_09_23_print.pdf). [Accessed 06-12-2024].