

A meshfree approach for rapid reconstruction of compact binary sources detected in the network of international gravitational wave detectors

A Thesis submitted
in partial fulfillment of the requirements for the Degree of

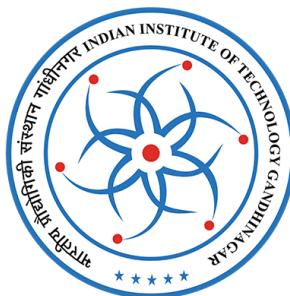
Doctor of Philosophy

by

Lalit Pathak

Roll No: 18310022

Indian Institute of Technology Gandhinagar



Under the guidance of

Prof. Anand Sengupta

Associate Professor

Department of Physics

Indian Institute of Technology Gandhinagar

Gandhinagar - 382055, Gujarat, India

June 2023

This page is intentionally left blank.

Certificate

It is certified that the work contained in the thesis titled "**A meshfree approach for rapid reconstruction of compact binary sources detected in the network of international gravitational wave detectors**" by "**Mr. Lalit Pathak**" (Roll no. 18310022), has been carried out under my supervision and this work has not been submitted elsewhere for a degree.

(Thesis Supervisor)
Prof. Anand Sengupta
Associate Professor
Department of Physics
Indian Institute of Technology Gandhinagar,
Gujarat, India

Acknowledgements

I extend my heartfelt appreciation to Professor Anand Sengupta for his unwavering guidance and thoughtful support. I am deeply grateful to my family for their constant love, care, and enduring support, demonstrating immense patience throughout my life's journey. Additionally, I am indebted to my friends at IIT Gandhinagar, whose continuous support and timely assistance have been invaluable. Special thanks go to Soumen Roy, Amit Reza, Rahul Shastri, Abhishek Sharma, Sanket Munishwar, and Sachin Shukla for their consistent encouragement and engaging discussions. I am grateful to the TCS Research Scholar program for providing the funding for my Ph.D. studies. I express my gratitude to IIT Gandhinagar for offering the awareness and resources that have significantly contributed to my personal and professional development throughout the duration of my five-and-a-half-year Ph.D. program.

Dedicated to my family and my love “Rashmi”

Contents

Publications	7
1 Introduction	9
1.1 Introduction to Gravitational waves (GWs)	9
1.1.1 Einstein’s equation and plane wave solution	9
1.1.2 Test masses interaction with GWs	13
1.1.3 GW interaction with detector	17
1.1.4 GW generation in linearized theory	22
1.1.5 Inspiring compact binaries	26
1.1.6 Indirect evidence and direct evidence of GWs	29
1.1.7 Detection strategies	34
1.1.8 GW Detectors: Interferometers	41
1.2 Parameter estimation	54
1.2.1 Inverse problems	54
1.2.2 Bayesian analysis	56
1.2.3 Sampling	57
1.3 Appendix	71
2 Fast and faithful interpolation of numerical relativity surrogate waveforms using meshfree approximation	73
2.1 Introduction	74
2.2 NR Waveform model	75
2.3 Generation of meshfree interpolants	78
2.3.1 Selection of patch(es)	78
2.3.2 Placing RBF nodes using ‘Iterative SVD’ method	79
2.4 Choice of RBF kernel	82
2.5 Results	84
2.6 Conclusion and Future Outlook	87
3 Fast likelihood evaluation using meshfree approximations for reconstructing compact binary sources	90
3.1 Introduction	90
3.2 Bayesian inference	92

3.3	Likelihood interpolation	93
3.3.1	Grid-based likelihood interpolation	94
3.3.2	Meshfree likelihood interpolation	97
3.4	Numerical experiments	99
3.4.1	2D likelihood interpolation	99
3.4.2	4D likelihood interpolation	101
3.5	Speed-up analysis	103
3.6	Conclusion and Outlook	105
3.7	Appendix	106
3.7.1	Accuracy trade-off: Effect of varying basis size	106
3.7.2	Scaling of the speed-up factor with waveform duration	106
3.7.3	RBF kernels and meshfree interpolation	108
3.7.4	Chebyshev polynomials	109
4	Prompt sky localization of compact binary sources using meshfree approximation	110
4.1	Introduction	110
4.2	Parameter estimation	113
4.2.1	Bayesian inference	113
4.2.2	The likelihood function	114
4.3	Meshfree likelihood interpolation	115
4.3.1	Start-up stage	116
4.3.2	Online stage	120
4.4	Numerical experiments	120
4.4.1	GW170817 event	120
4.4.2	Simulated Data	125
4.5	Conclusion and Outlook	127
4.6	Appendix	130
4.6.1	Effect of different choices of Gaussian distributed nodes on the posteriors	130
4.6.2	Likelihood errors	131
4.6.3	Surface of SVD coefficients and template norm square	131
4.6.4	Addition of tidal deformability in the meshfree framework	132
4.6.5	GW200115: An NSBH event	133
5	Pinpointing coalescing binary neutron star sources with the IGWN, including LIGO-Aundha	134
5.1	Introduction	135
5.1.1	Detector Networks	138
5.1.2	Duty Cycles	139
5.1.3	Injected BNS sources	140
5.1.4	Bayesian inference	141
5.2	Analysis of Simulated Events	145

5.3	Sky Localization Results	146
5.3.1	A1 at aLIGO-O4 sensitivity	149
5.3.2	A1 at A+ Design (O5) sensitivity	150
5.4	Localization of simulated BNS events subthreshold in A1	151
5.5	Experiments with GWTC-like events in real noise	152
5.5.1	Noise Strain and PSDs	152
5.5.2	Choice of injection parameters:	153
5.5.3	Analysis and Configurations	153
5.5.4	Degeneracy between luminosity distance and inclination angle	155
5.6	Conclusion	157
5.7	Discussion	158
5.8	Appendix	160
5.8.1	GWTC-like events corner plots	160
5.8.2	Simulated BNS event	161
6	Conclusion and Future Outlook	164

Publications

The chapters of this thesis are based on the following publications:

1. Fast and faithful interpolation of numerical relativity surrogate waveforms using meshfree approximation, **Lalit Pathak**, Amit Reza, Anand Sengupta, arXiv:2403.19162 (2024)
2. Fast likelihood evaluation using meshfree approximations for reconstructing compact binary sources, **Lalit Pathak**, Amit Reza, and Anand S. Sengupta, Phys. Rev. D **108**, 064055 (2023)
3. Prompt sky localization of compact binary coalescences using meshfree approximation, **Lalit Pathak**, Sanket Munishwar, Amit Reza, and Anand S. Sengupta, Phys. Rev. D **109**, 024053 (2024)
4. Pinpointing Binary Neutron Star sources with a global network of GW Detectors, including LIGO-Aundha, Sachin Shukla, **Lalit Pathak**, Anand Sengupta, Phys. Rev. D **109**, 044051 (2024)

Other publications not included in this thesis

1. Does the speed of gravitational waves depend on the source velocity?, Rajes Ghosh, Sreejith Nair, **Lalit Pathak**, Sudipta Sarkar, Anand S. Sengupta, Phys. Rev. D **108**, 124017 (2023)

Chapter 1

Introduction

1.1 Introduction to Gravitational waves (GWs)

1.1.1 Einstein's equation and plane wave solution

The detection of the first Gravitational wave signal in September 2015 [1, 2] ushered in a new era of gravitational wave astronomy. Albert Einstein, in his seminal work on General relativity (GR) [3], predicted the existence of gravitational waves [4, 5] but also lamented our possible inability to detect extremely faint signals using the technology available during his time. This raises the question of why Einstein held such reservations and what precisely we measure in a gravitational wave (GW) detector. To address such questions, we need to establish a theoretical foundation for these detections, beginning with the fundamental principles. Let's start with Einstein's field equations.

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu} \quad (1.1)$$

where $R_{\mu\nu}$, R , $g_{\mu\nu}$, and $T_{\mu\nu}$ are Ricci tensor, Ricci scalar, metric, and Energy-momentum tensor, respectively (For definitions, refer to Appendix 1.3). To see how GWs appear from Einstein's equations, let's linearize the field equations by considering a perturbation $h_{\mu\nu}$ on a flat Minkowski background $\eta_{\mu\nu}$

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad |h_{\mu\nu}| \ll 1, \quad (1.2)$$

Now, let's perform a finite, global Lorentz transformation defined by

$$x^\mu \rightarrow \Lambda_\nu^\mu x^\nu \quad (1.3)$$

Under Lorentz transformation, the matrix Λ_ν^μ transforms as

$$\eta_{\mu\nu} = \Lambda_\mu^\rho \Lambda_\nu^\sigma \eta_{\rho\sigma}, \quad (1.4)$$

which leads to the following transformation for $g_{\mu\nu}$:

$$\begin{aligned} g_{\mu\nu}(x) &\rightarrow g'_{\mu\nu}(x') = \Lambda_\mu^\rho \Lambda_\nu^\sigma \eta_{\rho\sigma} g_{\rho\sigma}(x) \\ &= \Lambda_\mu^\rho \Lambda_\nu^\sigma \eta_{\rho\sigma} [\eta_{\rho\sigma} + h_{\rho\sigma}(x)] \\ &= \eta_{\mu\nu} + \Lambda_\mu^\rho \Lambda_\nu^\sigma \eta_{\rho\sigma} h_{\rho\sigma}(x) = \eta_{\mu\nu} + h'_{\mu\nu}(x'), \end{aligned} \quad (1.5)$$

The above equation implies that $h_{\mu\nu}$ transforms as a tensor under Lorentz transformations. To linearize the Eq. (1.1), we need to linearize both $R_{\mu\nu}$ and R . First we express $R_{\mu\nu}$ in terms of Christoffel symbol $\Gamma_{\mu\nu}^\alpha$ as the following:

$$R_{\mu\nu} = R_{\mu\alpha\nu}^\alpha = \partial_\alpha \Gamma_{\mu\nu}^\alpha - \partial_\nu \Gamma_{\mu\alpha}^\alpha + \Gamma_{\beta\alpha}^\alpha \Gamma_{\mu\nu}^\beta - \Gamma_{\beta\nu}^\alpha \Gamma_{\mu\alpha}^\beta \quad (1.6)$$

where the Christoffel symbol is given by

$$\Gamma_{\mu\nu}^\alpha = \frac{1}{2} g^{\alpha\sigma} (\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}) \quad (1.7)$$

The inverse metric $g^{\mu\nu}$ can be written as $\eta^{\mu\nu} - h^{\mu\alpha}$. Pluggin it in Eq. (1.7) along with Eq. (1.2), to $\mathcal{O}(h)$, we find

$$\Gamma_{\mu\nu}^\alpha = \frac{1}{2} \eta^{\alpha\sigma} (\partial_\mu h_{\sigma\nu} + \partial_\nu h_{\sigma\mu} - \partial_\sigma h_{\mu\nu}) \quad (1.8)$$

Pluggin the above expression in Eq. (1.6), the linearized Riemann tensor to $\mathcal{O}(h)$ becomes

$$R_{\mu\nu\rho\sigma} = \frac{1}{2} (\partial_\nu \partial_\rho h_{\mu\sigma} + \partial_\mu \partial_\sigma h_{\nu\rho} - \partial_\nu \partial_\sigma h_{\mu\rho}) \quad (1.9)$$

To compactify the linearized equations, we define

$$h = \eta^{\mu\nu} h_{\mu\nu} \quad (1.10)$$

and

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} h \quad (1.11)$$

From Eq. (1.11), using $\bar{h} = -h$, and inverting the above equation, we find

$$h_{\mu\nu} = \bar{h}_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} \bar{h} \quad (1.12)$$

Using Eq. (1.9) and Eq. (1.12), the linearization of Einstein's equations, Eq. (1.1) leads to

$$\square \bar{h}_{\mu\nu} + \eta_{\mu\nu} \partial^\rho \partial^\sigma \bar{h}_{\rho\sigma} - \partial^\rho \partial_\nu \bar{h}_{\mu\rho} - \partial^\rho \partial_\mu \bar{h}_{\nu\rho} = -\frac{16\pi G}{c^4} T_{\mu\nu} \quad (1.13)$$

where $\square \equiv -\frac{1}{c^2} \partial_0^2 + \nabla^2$.

Now, recalling that we chose a frame where Eq. (1.2) is valid, there is still a residual gauge symmetry that can be exploited. Under a coordinate transformation defined by $x^\mu \rightarrow x'^\mu = x^\mu + \psi^\mu(x)$ and using transformation law (Eq. (1.5)), we find

$$h'_{\mu\nu}(x') = h_{\mu\nu}(x) - (\partial_\sigma \psi_\rho + \partial_\rho \psi_\sigma) - (\partial_\sigma \psi^\mu h'_{\mu\sigma}(x') + \partial_\sigma \psi^\nu h'_{\rho\nu}(x')) \quad (1.14)$$

Assuming $|\partial_\mu \psi_\nu| \sim \mathcal{O}(|h_{\mu\nu}|)$, the above expression in linear order becomes

$$h'_{\mu\nu}(x') = h_{\mu\nu}(x) - (\partial_\sigma \psi_\rho + \partial_\rho \psi_\sigma) \quad (1.15)$$

Using the above transformation rule, we find

$$\begin{aligned} \bar{h}'_{\mu\nu} &= h_{\mu\nu} - (\partial_\mu \psi_\nu + \partial_\nu \psi_\mu) - \frac{1}{2} \eta_{\mu\nu} h + \eta_{\mu\nu} \eta^{\alpha\beta} \partial_\alpha \psi_\beta \\ &= \bar{h}_{\mu\nu} - (\partial_\mu \psi_\nu + \partial_\nu \psi_\mu + \eta_{\mu\nu} \partial_\rho \psi^\rho) \end{aligned} \quad (1.16)$$

which implies

$$(\partial^\nu \bar{h}_{\mu\nu})' = \partial^\nu \bar{h}_{\mu\nu} - \square \psi_\mu \quad (1.17)$$

Now, we can always choose $\square \psi_\mu$ such that $(\partial^\nu \bar{h}_{\mu\nu})' = 0$. It is known by the names of Lorentz/Hilbert/harmonic/De Donder gauge. In this gauge, the last three terms of the Eq. (1.13) vanishes, giving us the following wave equation:

$$\square \bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4} T_{\mu\nu} \quad (1.18)$$

Note that we started with a symmetric $h_{\mu\nu}$ with **ten** independent components. With the Lorentz gauge, we have 4 conditions, which leads $h_{\mu\nu}$ to **six** independent components. To understand the propagation of the GWs and their effect on a GW detector, which is essentially happening outside the source, we can put $T_{\mu\nu} = 0$. It implies that

$$\square \bar{h}_{\mu\nu} = 0 \quad (1.19)$$

Since the Lorentz gauge didn't fix the gauge completely, let's perform another coordinate transformation such that $\square \psi_\mu = 0$. It also implies $\square \psi_{\mu\nu} = 0$, where

$$\psi_{\mu\nu} \equiv \partial_\mu \psi_\nu + \partial_\nu \psi_\mu - \eta_{\mu\nu} \partial_\rho \psi^\rho, \quad (1.20)$$

So we now have four extra conditions, coming from $\square \psi_{\mu\nu} = 0$, since it depends on four functions ψ_μ . These conditions can be imposed on $\bar{h}_{\mu\nu}$, which already have six independent components. Furthermore, ψ^0 can be chosen such that $\bar{h} = 0$ which also implies that $\bar{h}_{\mu\nu} = h_{\mu\nu}$. The rest of the functions $\psi^i(x)$ are chosen such that $h^{0i}(x) = 0$. This is what is known as the transverse-traceless gauge (TT gauge). So, finally, we have reduced the **ten** degrees of freedom associated with $h_{\mu\nu}$ to **two** degrees of freedom by exploiting the Lorentz gauge and a residual gauge associated with four functions ψ^μ . Unless otherwise stated, we will refer to metric in TT gauge as $h_{\mu\nu}$.

The Eq. (1.19) have plane-wave solutions, $h_{ij}(x) = e_{ij}(\vec{k}) e^{jkx}$ with four vector $k^\mu = (\omega/c, \vec{k})$ and $\omega/c = |\vec{k}|$, where k^μ has dimensions of inverse length, and $e_{ij}(\vec{k})$ is the polarization tensor. From the above discussion, it is clear that the non-zero components of h_{ij} lie in a plane perpendicular to \hat{n} and the condition $\partial^j h_{ij} = 0$ is equivalent to $n^i h_{ij} = 0$ for a plane wave. With $\hat{n} = \vec{k}/|\vec{k}|$ (direction of wave propagation) along the z -axis and a symmetric and traceless h_{ij} , we get

$$h_{ij}(t, z) = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}_{ij} \cos[\omega(t - z/c)], \quad (1.21)$$

In particular, we can simplify the above expression as the following:

$$h_{ab}(t, z) = \begin{pmatrix} h_+ & h_\times \\ h_\times & -h_+ \end{pmatrix}_{ab} \cos[\omega(t - z/c)], \quad (1.22)$$

where h_+ and h_\times are called the amplitudes of the “plus” and “cross” polarization of the wave, and indices $a, b = 1, 2$ represent the non-zero components of h_{ij} . The equation of motion Eq. (1.19) implies that we can expand h_{ij} as

$$h_{ij}(x) = \int \frac{d^3 k}{(2\pi)^3} (\mathcal{B}_{ij}(\vec{k}) e^{jkx} + \mathcal{B}_{ij}^*(\vec{k}) e^{-jkx}) \quad (1.23)$$

where $d^3k = (2\pi/c)^3 f^2 df d\Omega$ with $f > 0$. Denoting $d^2\hat{n} = d\cos\theta d\theta d\phi$ as the integration over solid-angle, the above equation reads,

$$h_{ij}(x) = \frac{1}{c^3} \int_0^\infty df f^2 \int d^2\hat{n} (\mathcal{B}_{ij}(f, \hat{n}) e^{-2\pi j f(t - \hat{n} \cdot x/c)} + c.c.) \quad (1.24)$$

The TT gauge conditions implies that $\mathcal{B}_i^i(\vec{k})$ and $\mathcal{B}_{ij}(\vec{k}) = 0$. Note that for superimposition of waves with different \hat{n} , h_{ij} will no longer be a 2×2 matrix. However, for the GWs detected on the Earth, \hat{n} is very well defined and therefore, can be written as

$$\mathcal{B}_{ij}(\vec{k}) = B_{ij}(f) \delta^{(2)}(\hat{n} - \hat{n}_0) \quad (1.25)$$

where \hat{n}_0 is the direction of wave propagation.

Using Eq. (1.24), we can write

$$h_{ab}(t, \vec{x}) = \int_0^\infty df \left(\tilde{h}_{ab}(f, \vec{x}) e^{-2\pi j f t} + \tilde{h}_{ab}^*(f, \vec{x}) e^{2\pi j f t} \right), \quad (1.26)$$

where

$$\begin{aligned} \tilde{h}_{ab}(f, \vec{x}) &= \frac{f^2}{c^3} \int d^2\hat{n} \mathcal{A}_{ab}(f, \hat{n}) e^{2\pi j f \hat{n} \cdot x/c} \\ &= \frac{f^2}{c^3} A_{ab}(f) e^{2\pi j f \hat{n}_0 \cdot x/c} \end{aligned} \quad (1.27)$$

From Eq. (1.22) and the above equation, it follows that

$$\tilde{h}_{ab}(f) = \begin{pmatrix} \tilde{h}_+(f) & \tilde{h}_\times(f) \\ \tilde{h}_\times(f) & -\tilde{h}_+(f) \end{pmatrix}_{ab} \quad (1.28)$$

where the “+” and “×” polarizations are defined with respect to a given choice of axes in the transverse plane. As evident from Eq. (1.26), only physical frequencies ($f > 0$) are present in the equations. We can extend it to negative frequencies by defining

$$\tilde{h}_{ab}(-f, \vec{x}) = \tilde{h}_{ab}^*(f, \vec{x}), \quad (1.29)$$

plugging in the Eq. (1.26), we get

$$h_{ab}(t) = \int_{-\infty}^\infty df \tilde{h}_{ab}(f) e^{-2\pi j f t}. \quad (1.30)$$

Apart from the above forms, we can express the $h_{ab}(t, \vec{x})$ in terms of so called polarization tensors $e_{ij}^\alpha(\hat{n})$ (with α defining labels + and ×) which are defined as

$$\begin{aligned} e_{ij}^+(\hat{n}) &= \hat{u}_i \hat{u}_j - \hat{v}_i \hat{v}_j \\ e_{ij}^\times(\hat{n}) &= \hat{u}_i \hat{v}_j + \hat{v}_i \hat{u}_j, \end{aligned} \quad (1.31)$$

where \hat{u} and \hat{v} are unit vectors orthogonal to \hat{n} and to each other. With \hat{n} along z -axis in a given frame, one can choose $\hat{u} = \hat{x}$ and $\hat{v} = \hat{y}$ and express polarization tensors in the following form:

$$\begin{aligned} e_{ab}^+ &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}_{ab} \\ e_{ab}^\times &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}_{ab} \end{aligned} \quad (1.32)$$

In a generic frame, the amplitudes $\tilde{h}_A(f, \hat{n})$ can be read from the following:

$$\frac{f^2}{c^3} \mathcal{A}_{ij}(f, \hat{n}) = \sum_{\alpha=+, \times} \hat{h}_\alpha(f, \hat{n}) e_{ij}^\alpha(\hat{n}) \quad (1.33)$$

The Eq. (1.24) now be written as

$$h_{ab}(t, \vec{x}) = \sum_{\alpha=+, \times} \int_{-\infty}^{\infty} df \int d^2 \hat{n} \hat{h}_\alpha(f, \hat{n}) e_{ab}^\alpha(\hat{n}) e^{-2\pi j f(t - \hat{n} \cdot \vec{x}/c)}, \quad (1.34)$$

with $\hat{h}_\alpha(-f, \hat{n}) = \tilde{h}_\alpha^*(f, \hat{n})$. Now, we will discuss how the GWs interact with test masses.

1.1.2 Test masses interaction with GWs

Geodesic equation and Local inertial frame

In GR, the choice of a particular gauge is synonymous with the selection of a specific observer. As discussed in the preceding section, GWs exhibit a simplified expression in the TT gauge. This prompts the question of what reference frame corresponds to the TT gauge. In order to understand the effects of GWs on test masses, let's start with a geodesic equation which is given by

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma_{\nu\rho}^\mu(x) \frac{dx^\nu}{d\tau} \frac{dx^\rho}{d\tau} = 0 \quad (1.35)$$

which describes the motion of a test mass in curved background metric $g_{\mu\nu}$ in the absence of non-gravitational forces, and where τ is the proper time. The above equation can rewritten in terms of four-velocity u^μ as

$$\frac{du^\mu}{d\tau} + \Gamma_{\nu\rho}^\mu u^\nu u^\rho = 0 \quad (1.36)$$

To find the equation of geodesic deviation, let's consider two nearby geodesics represented by $x^\mu(\tau)$ and the $x^\mu(\tau) + \zeta^\mu(\tau)$. The geodesic equation for the latter is given by

$$\frac{d^2(x^\mu + \zeta^\mu)}{d\tau^2} + \Gamma_{\nu\rho}^\mu(x + \zeta) \frac{d(x^\nu + \zeta^\nu)}{d\tau} \frac{d(x^\rho + \zeta^\rho)}{d\tau} = 0 \quad (1.37)$$

If $|\zeta^\mu|$ is much smaller than the typical scale of variation of the gravitational field, the above equation can be expanded to first order in ζ as

$$\frac{d^2 \zeta^\mu}{d\tau^2} + 2 \Gamma_{\nu\rho}^\mu(x) \frac{dx^\nu}{d\tau} \frac{d\zeta^\rho}{d\tau} + \zeta^\sigma \partial_\sigma \Gamma_{\nu\rho}^\mu(x) \frac{dx^\nu}{d\tau} \frac{dx^\rho}{d\tau} \quad (1.38)$$

It can be written in a more elegant way in terms of the covariant derivative (for definition, refer to Eq. (1.244) in Appendix 1.3) of a vector field ζ^μ along the curve $x^\mu(\tau)$ as

$$\frac{D^2 \zeta^\mu}{d\tau^2} = -R_{\nu\rho\sigma}^\mu \zeta^\rho \frac{dx^\nu}{d\tau} \frac{x^\sigma}{d\tau}, \quad (1.39)$$

The implication of the equation mentioned above is that a tidal gravitational force exists between nearby time-like geodesics, and its determination relies on the Riemann tensor. This understanding aids in elucidating the behavior of test masses influenced by gravitational waves for a specific observer. Before we delve further into a crucial discussion on reference frames, let us briefly discuss the construction of a local inertial frame (LIF). Within the framework of

General Relativity (GR), it is always possible to perform coordinate transformations in such a manner that, at a specific spacetime point P, all components of the Christoffel symbol are zero identically ($\Gamma_{\nu\rho}^\mu(P) = 0$). This implies that the test mass is in a state of free fall in this reference frame, albeit only at a particular point in space and time.

$$\frac{d^2x^\mu}{d\tau^2}\Big|_P = 0 \quad (1.40)$$

As previously explained, within the Local Inertial Frame (LIF), the test mass moves freely at a specific point in spacetime. Alternatively, we can construct a reference frame in which the test mass remains in free fall throughout its journey along the geodesic. This can be achieved by employing a freely spinning gyroscope in motion along a time-like geodesic $x^\mu(\tau)$ obeying the following equation:

$$\frac{ds^\mu}{d\tau} + \Gamma_{\nu\rho}^\mu s^\nu \frac{dx^\rho}{d\tau} = 0 \quad (1.41)$$

where s^μ is the spin four-vector. We can start with an LIF at P with three gyroscopes defining the directions of the spatial axes. During the movement along the geodesic, the spatial axes continuously orient themselves with the directions indicated by gyroscopes. Note that since gyroscopes define the orientation of the axes, they do not rotate relative to the axes within this frame. This results in $ds^\mu/d\tau = 0$ throughout the geodesic. In line with Eq. (1.41), this condition implies that $\Gamma_{\nu\rho}^\mu$ vanishes along the entire geodesic. Such a reference frame is commonly referred to as a freely-falling frame. For a more comprehensive and detailed discussion of the construction of the Local Inertial Frame, please refer to Section 8.4 of Hartle [6].

TT frame and proper detector frame

As found in the earlier sections, in the TT gauge, the GWs have a simple form. We now seek to determine the reference frame (or observer) associated with this gauge. Writing down the Eq. (1.35) for a test mass at rest at $\tau = 0$

$$\begin{aligned} \frac{d^2x^i}{d\tau^2}\Big|_{\tau=0} &= - \left[\Gamma_{\nu\rho}^i(x) \frac{dx^\nu}{d\tau} \frac{dx^\rho}{d\tau} \right] \Big|_{\tau=0} \\ &= - \left[\Gamma_{00}^i \left(\frac{dx^0}{d\tau} \right)^2 \right]_{\tau=0}, \end{aligned} \quad (1.42)$$

From Eq. (1.8), we find that

$$\Gamma_{00}^i = \frac{1}{2}(2\partial_0 h_{0i} - \partial_i h_{00}). \quad (1.43)$$

Since $h_{00} = h_{0i} = 0$ in the TT gauge, the above equation implies $d^2x^i/d\tau^2 = 0$ at $\tau = 0$. It means that the test mass remains at rest at all times, implying that the GWs have no effect on the motion of the test mass. To see what happens to the separation of two test masses which are initial at rest, the spatial component of the Eq. (1.38) is written as

$$\frac{d^2\xi^i}{d\tau^2}\Big|_{\tau=0} = - \left[2c\Gamma_{0\rho}^i \frac{d\xi^\rho}{d\tau} + c^2 \xi^\sigma \partial_\sigma \Gamma_{00}^i \right]_{\tau=0} \quad (1.44)$$

The first term in the above equation is non-zero only when ρ represents a spatial index, whereas the second term is zero identically. From Eq. (1.8), we find $\Gamma_{0j}^i = (\frac{1}{2})\partial_0 h_{ij}$. Therefore, the above equation is reduced to

$$\frac{d^2\zeta^i}{d\tau^2}\Big|_{\tau=0} = - \left[\dot{h}_{ij} \frac{d\zeta^i}{d\tau} \right]_{\tau=0}, \quad (1.45)$$

which implies that if at $\tau = 0$, the $d\zeta^i/d\tau = 0$, then $\frac{d^2\zeta^i}{d\tau^2} = 0$ and hence the separation ζ^i remains constant at all times.

From the above discussion, it seems as if there is no physical effect of GWs passing by test masses. However, the physical effects of GWs indeed exist and can be found by monitoring the proper distances instead of coordinate distances. To see this, let us write the proper distance between two events (say $(t, x_1, 0, 0)$ and $(t, x_2, 0, 0)$) as

$$\begin{aligned} s &= (x_2 - x_1)[1 + h_+ \cos(\omega t)]^{1/2} \\ &\simeq L [1 + \frac{1}{2}h_+ \cos(\omega t)], \end{aligned} \quad (1.46)$$

which implies that the proper distance changes periodically in time due to GWs. Generally, for two events separated by a vector \vec{L} , the proper distance is given by $s^2 = L^2 + h_{ij}(t)L_iL_j$, where $h_{ij}(t)$ is the GW tensor at time t . To linear order in h , this equation can be approximated as $s \simeq L + h_{ij}(L_iL_j/2L)$, giving

$$\begin{aligned} \ddot{s} &\simeq \frac{1}{2}\ddot{h}_{ij} \frac{L_i}{L}L_j \\ &\simeq \frac{1}{2}\ddot{h}_{ij}s_j, \end{aligned} \quad (1.47)$$

where $n_i = L_i/L$ and $s = n$. This is the geodesic equation expressed in terms of proper distances. The fact that GWs affect proper distances implies that any quantity dependent on proper distances can be measured, facilitating the detection of GWs. In fact, this forms the basis of interferometer-based GW detectors, where a laser beam travels back and forth between the mirrors (akin to test masses) whose round trip time is dependent on the proper distance. We will study this in the later sections while discussing the GW detectors. Now, let's briefly introduce the proper detector frame. As mentioned earlier, the TT frame simplifies the form of GWs. However, it may not be a suitable frame for experimentalists describing GW detectors since, in a detector, the positions are not represented by freely falling particles. It can be demonstrated that in the proper detector frame, the equation of geodesic deviation¹ can be expressed as follows:

$$\ddot{\xi}^i = \frac{1}{2}\ddot{h}_{ij}^{\text{TT}}\xi^j \quad (1.48)$$

Unlike Eq. (1.45), the above equation implies that the effects of GW on a point particle of mass m can be expressed in terms of a Newtonian force in the proper detector frame as

$$F_i = \frac{m}{2}\ddot{h}_{ij}^{\text{TT}}\xi^j, \quad (1.49)$$

¹For derivation, refer [7]

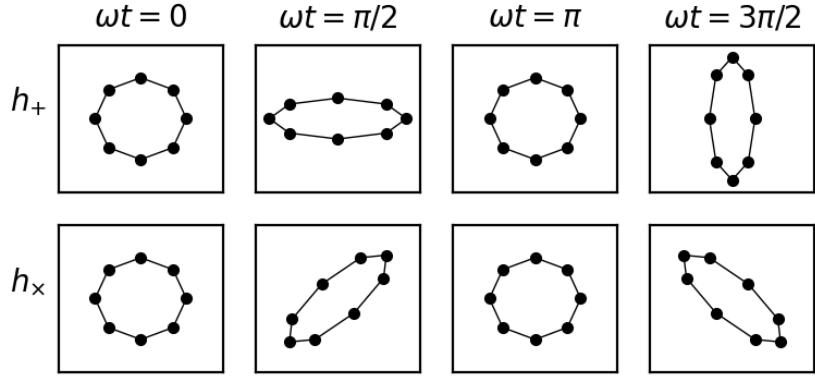


Figure 1.1: The ring of test masses shows deformation due + and \times polarizations of GWs as they enter into the plane of the paper. Notice that the pattern ring follows as the phase changes. h_+ polarization follows a “+” pattern while h_x follows a “ \times ” pattern and hence the denominations “plus” and “cross” polarizations.

which is a great simplification over the TT frame. Naturally, there are other slowly varying Newtonian forces, the influence of which can be minimized by carefully selecting a frequency window $[f_{min}, f_{max}]$. Here, the choice of f_{min} ensures that the slowly varying Newtonian and seismic noise remains significantly low, while f_{max} is determined to prevent other sources of noise from dominating. Consequently, the proper detector frame offers two key advantages: (i) in this frame, the description of the detector can be done in Newtonian language, which is more intuitive, and (ii) The Newtonian force induced by gravitational waves can be expressed in terms of h_{ij} , which is calculated in the TT gauge and has a highly simplified form. Now, let us examine the effect of GWs on a ring of test masses which is placed in the x - y plane and a GW is propagating along the z -direction. In such a scenario, the motion of the test particles remains confined to the x - y plane in accordance with Eq. (1.48). To begin, let’s consider the plus h_+ polarization, which at $z = 0$ can be represented as:

$$h_{ab} = h_+ \sin(\omega t) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (1.50)$$

We choose $\zeta_a(t)$ (see Eq. (1.48)) to be $(x_0 + \delta x(t), y_0 + \delta y(t))$, where x_0 and y_0 are the unperturbed positions and $\delta x(t)$ and $\delta y(t)$ induced by the GW. From Eq. (1.48), we find

$$\delta \ddot{x} = -\frac{h_+}{2} (x_0 + \delta x) \omega^2 \sin(\omega t), \quad (1.51a)$$

$$\delta \ddot{y} = \frac{h_+}{2} (y_0 + \delta y) \omega^2 \sin(\omega t). \quad (1.51b)$$

Neglecting the $O(h)$ in the right-hand side of the above equations and integrating, we get

$$\delta x = \frac{h_+}{2} x_0 \sin(\omega t), \quad (1.52a)$$

$$\delta y = -\frac{h_+}{2} y_0 \sin(\omega t). \quad (1.52b)$$

Similarly for cross h_{\times} polarizations, we find

$$\delta x = \frac{h_{\times}}{2} y_0 \sin(\omega t), \quad (1.53a)$$

$$\delta y = \frac{h_{\times}}{2} x_0 \sin(\omega t). \quad (1.53b)$$

Fig. 1.1 shows how a ring of test masses is deformed by a passing gravitational wave (GW). We now briefly discuss how the detector responds to GWs.

1.1.3 GW interaction with detector

We again revisit the end of Sec. 1.1.1, and we saw that we could describe the GWs in the TT gauge as a linear combination of polarization tensors e_{ij}^{α} .

$$h_{ij} = h_{+} e_{ij}^{+} + h_{\times} e_{ij}^{\times} \quad (1.54a)$$

$$\mathbf{h} = h_{+} \mathbf{e}^{+} + h_{\times} \mathbf{e}^{\times} \quad (1.54b)$$

where $\{e_{ij}^{+}\}$ and $\{e_{ij}^{\times}\}$ can be written as

$$\{e_{ij}^{+}\} \equiv \mathbf{e}^{+} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \{e_{ij}^{\times}\} \equiv \mathbf{e}^{\times} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (1.55)$$

However, it is more useful to express these quantities in a coordinate-independent manner. For example, we can represent a vector \vec{A} in different bases, e.g., $\{\vec{e}_i\}$ and $\{\vec{e}'_i\}$ and its components in these coordinate systems are defined by

$$A_i = \vec{e}_i \cdot \vec{A} \quad \text{or} \quad A'_i = \vec{e}'_i \cdot \vec{A} \quad (1.56)$$

We can represent the same physical quantity in various manners without altering the underlying physics. To achieve this, we will introduce an abstract index notation, where A_a is equivalent to the familiar vector \vec{A} , referring to an object with components $\{A_i\}$ when analyzed within the basis \vec{e}_i . Returning to Eq. (1.54), we would like to formulate it in a manner that is independent of a specific set of basis vectors. This can be accomplished by constructing matrices \mathbf{e}^{+} and \mathbf{e}^{\times} using the components of the unit vectors $\vec{\ell}$ and \vec{m} . These unit vectors, in combination with \vec{k} —representing the direction of gravitational wave propagation—form an orthonormal triple. In the coordinate system we have been working with, these vectors have the following components:

$$\{\ell_i\} = \boldsymbol{\ell} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \{m_i\} = \boldsymbol{m} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad (1.57)$$

and we can express \mathbf{e}^{+} and \mathbf{e}^{\times} in terms of the above vectors as

$$\mathbf{e}^{+} = \boldsymbol{\ell} \boldsymbol{\ell}^T - \boldsymbol{m} \boldsymbol{m}^T \quad (1.58a)$$

$$\mathbf{e}^{\times} = \boldsymbol{\ell} \boldsymbol{m}^T + \boldsymbol{m} \boldsymbol{\ell}^T \quad (1.58b)$$

Or, in terms of components,

$$e_{ij}^+ = \ell_i \ell_j - m_i m_j \quad (1.59a)$$

$$e_{ij}^\times = \ell_i m_j + m_i \ell_j \quad (1.59b)$$

In the abstract index notation, we can describe the above polarization matrices in the following way:

$$e_{ab}^+ = \ell_a \ell_b - m_a m_b \quad (1.60a)$$

$$e_{ab}^\times = \ell_a m_b + m_a \ell_b \quad (1.60b)$$

Or, these basis vectors can be expressed in terms of the tensor (dyad) products as

$$\overset{\leftrightarrow}{e}^+ = \vec{\ell} \otimes \vec{\ell} - \vec{m} \otimes \vec{m} \quad (1.61a)$$

$$\overset{\leftrightarrow}{e}^\times = \vec{\ell} \otimes \vec{m} + \vec{m} \otimes \vec{\ell} \quad (1.61b)$$

using which we can write the general plane wave propagating along \vec{k} in covariant tensor notation as

$$\overset{\leftrightarrow}{h} = h_+ \overset{\leftrightarrow}{e}^+ + h_\times \overset{\leftrightarrow}{e}^\times \quad (1.62)$$

where h_+ and h_\times are the functions of $t - \vec{k} \cdot \vec{r}/c$. Note that the propagation direction alone

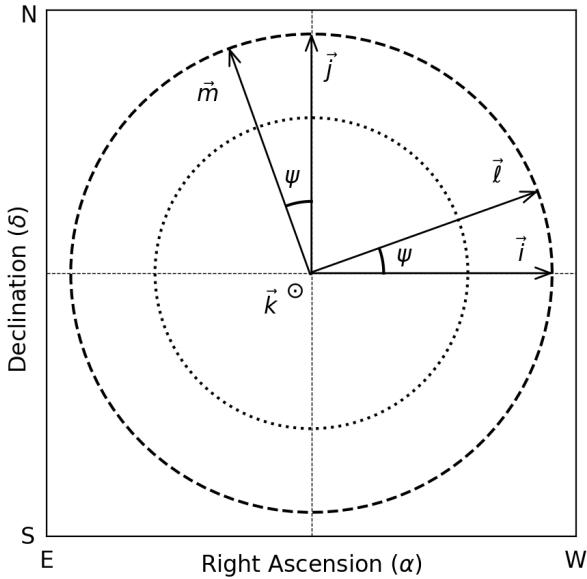


Figure 1.2: Natural polarization basis is obtained by rotating the reference basis around the direction of propagation of GW. \vec{i} is chosen in the direction of decreasing α , so that \vec{j} points into the Northern celestial hemisphere. (reproduced from [8])

is insufficient to uniquely determine the construction of $\overset{\leftrightarrow}{e}^+$ and $\overset{\leftrightarrow}{e}^\times$. To establish a complete orthonormal basis, the selection of a vector $\vec{\ell}$ within the plane perpendicular to \vec{k} is also necessary. This choice finalizes the definition of the orthonormal triple comprising $\vec{\ell}$, \vec{m} , and \vec{k} . The specific polarization basis adopted may be influenced by the nature of the sources being studied and the requirements of the analysis. In some cases, a “convenient” polarization basis may be preferred.

It might also be desirable to transition from this convenient polarization basis to a reference basis constructed solely based on the direction of GW propagation and some absolute reference directions. For a clearer illustration, let's consider an example where the source's position in the sky is specified using its right ascension α and declination δ , which is equivalent to defining the propagation direction \vec{k} . At each sky location, a reference basis can be assigned by introducing two additional unit vectors, \vec{i} and \vec{j} , which together form an orthonormal set alongside \vec{k} . From these unit vectors, a reference polarization basis can be constructed for traceless symmetric tensors oriented transversely to \vec{k} . This basis can be expressed as follows:

$$\overset{\leftrightarrow}{\varepsilon}^+ = \vec{i} \otimes \vec{i} - \vec{j} \otimes \vec{j} \quad (1.63a)$$

$$\overset{\leftrightarrow}{\varepsilon}^\times = \vec{i} \otimes \vec{j} + \vec{j} \otimes \vec{i} \quad (1.63b)$$

Moreover, $\vec{\ell}, \vec{m}$ forming the natural polarization basis, lie in the same plane as \vec{i} and \vec{j} and hence perpendicular to \vec{k} . It implies that we can go from the reference basis to the natural basis by rotating the reference basis counter-clockwise (\vec{i} to $\vec{\ell}$) by some angle ψ and can be written in the following form:

$$\vec{\ell} = \vec{i} \cos \psi + \vec{j} \sin \psi \quad (1.64a)$$

$$\vec{m} = -\vec{i} \sin \psi + \vec{j} \cos \psi \quad (1.64b)$$

Substituting the above equations in Eq. (1.61), we find,

$$\overset{\leftrightarrow}{e}^+ = \vec{i} \otimes \vec{i} \cos 2\psi + \vec{i} \otimes \vec{j} \sin 2\psi + \vec{j} \otimes \vec{i} \sin 2\psi - \vec{j} \otimes \vec{j} \cos 2\psi \quad (1.65a)$$

$$\overset{\leftrightarrow}{e}^\times = -\vec{i} \otimes \vec{i} \sin 2\psi + \vec{i} \otimes \vec{j} \cos 2\psi + \vec{j} \otimes \vec{i} \cos 2\psi + \vec{j} \otimes \vec{j} \sin 2\psi \quad (1.65b)$$

which can be expressed in terms of $\overset{\leftrightarrow}{\varepsilon}^+$ and $\overset{\leftrightarrow}{\varepsilon}^\times$ using the Eq. (1.63) as

$$\overset{\leftrightarrow}{e}^+ = \overset{\leftrightarrow}{\varepsilon}^+ \cos 2\psi + \overset{\leftrightarrow}{\varepsilon}^\times \sin 2\psi \quad (1.66a)$$

$$\overset{\leftrightarrow}{e}^\times = -\overset{\leftrightarrow}{\varepsilon}^+ \sin 2\psi + \overset{\leftrightarrow}{\varepsilon}^\times \cos 2\psi \quad (1.66b)$$

which shows that three angles are needed to specify a polarization basis associated with a particular source: (i) α , (ii) δ , and (iii) an additional angle ψ , which is also known as the polarization angle defining the orientation of the preferred polarization basis $\{\overset{\leftrightarrow}{e}^+, \overset{\leftrightarrow}{e}^\times\}$ relative to some reference basis $\{\overset{\leftrightarrow}{\varepsilon}^+, \overset{\leftrightarrow}{\varepsilon}^\times\}$.

Now, we discuss how GWs affect a GW detector. For the sake of simplicity, we will focus on an “L” shaped GW interferometer, where each arm has a fixed length L_0 . A natural question that arises is how we can make measurements when both the spacetime and the detector itself are subject to stretching and squeezing due to GWs. To clarify, we actually measure the phase difference between the light beams returning from the two arms of the interferometer after reflecting off the mirrors. This phase difference is essentially equivalent to determining the roundtrip travel time of photons traveling down both arms. The beauty of this measurement lies in the fact that it remains unaffected by the presence of gravitational waves, mainly because the time component of the spacetime metric remains constant in the TT gauge.

Additionally, in the TT gauge, points with constant coordinates experience free fall. Assuming the beam splitter is located at the origin and the end mirror of one of the arms is positioned at coordinates $(x^1, x^2, x^3) = (L_0, 0, 0)$, these coordinates remain unchanged despite the passage of GWs. Considering the path of the photon traveling from the origin $(0, 0, 0)$ to the end mirror at $(L_0, 0, 0)$ and back, its trajectory follows the equation:

$$ds^2 = -c^2 dt^2 + (1 + h_{11})(dx^1)^2 = 0 \quad (1.67)$$

which implies,

$$dt = \frac{\sqrt{1 + h_{11}}}{c} |dx^1| \approx \left(1 + \frac{1}{2}h_{11}\right) \frac{|dx^1|}{c} \quad (1.68)$$

In general the GW amplitude h_{11} is a function of $\left(t - \frac{\vec{k} \cdot \vec{r}}{c}\right)$. However, in the “long-wavelength” limit where the wavelength of the GW is much longer than the arm length of the detector, we can approximate h_{11} to be a constant along the photon’s travel path. This approximation is not valid for space-borne GW detectors, such as LISA [9, 10, 11]. Under this approximation, the time taken by a photon to travel from $(0, 0, 0)$ to $(L_0, 0, 0)$ and back is

$$T_1 = \left(1 + \frac{1}{2}h_{11}\right) \frac{2L_0}{c} \quad (1.69)$$

We now describe the detector arm in a coordinate-independent manner and define a unit vector \vec{u} along one of the arms whose components (in a specialized coordinate system) are

$$\{u^i\} \equiv \mathbf{u} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad (1.70)$$

The metric perturbation h_{11} above can be expressed as

$$h_{11} = u^i h_{ij} u^j = \mathbf{u}^T \mathbf{h} \mathbf{u} \quad (1.71)$$

which can be written in the abstract index notations as

$$u^a h_{ab} u^b = \vec{u} \cdot \overset{\leftrightarrow}{h} \cdot \vec{u} \quad (1.72)$$

so that

$$L_{\vec{u}} = L_0 \left(1 + \frac{1}{2}h_{ab} u^a u^b\right) = L_0 \left(1 + \frac{1}{2}\vec{u} \cdot \overset{\leftrightarrow}{h} \cdot \vec{u}\right) \quad (1.73)$$

Choosing a unit vector \vec{v} along the other arm, the difference in the roundtrip times down the two arms is given by

$$\frac{2(L_{\vec{u}} - L_{\vec{v}})}{c} = \frac{1}{c} L_0 \left(\vec{u} \cdot \overset{\leftrightarrow}{h} \cdot \vec{u} - \vec{v} \cdot \overset{\leftrightarrow}{h} \cdot \vec{v}\right) \quad (1.74)$$

Dividing the above equation by $2L_0/c$, what we find is the Gravitational wave strain $h(t)$:

$$\begin{aligned} h &= \frac{1}{2} \left(\vec{u} \cdot \overset{\leftrightarrow}{h} \cdot \vec{u} - \vec{v} \cdot \overset{\leftrightarrow}{h} \cdot \vec{v}\right) \\ &= h_{ab} \frac{u^a u^b - v^a v^b}{2} = h_{ab} D^{ab} = \overset{\leftrightarrow}{h} : \overset{\leftrightarrow}{D} \end{aligned} \quad (1.75)$$

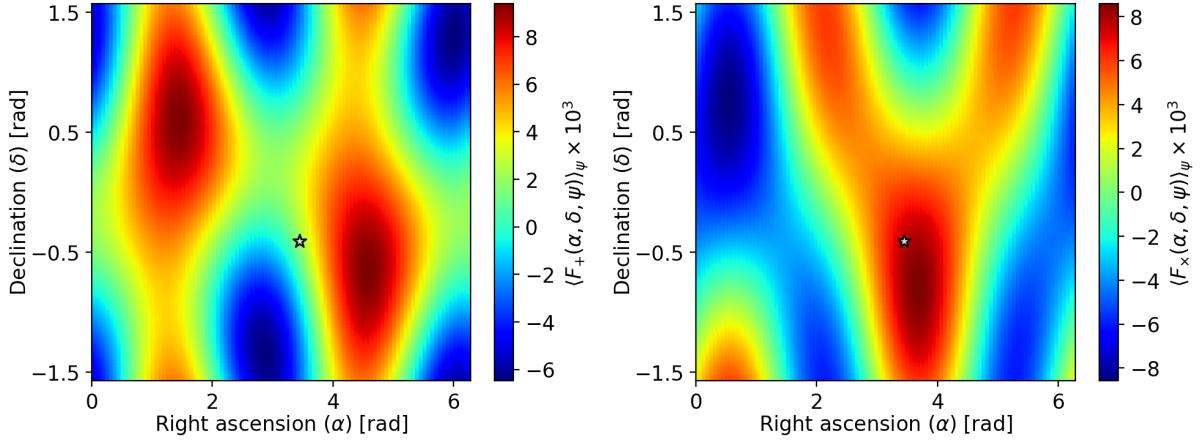


Figure 1.3: Antenna pattern functions F_+ and F_\times averaged over the polarization angle ψ , for a given detector (in this case LIGO Livingston “L1” [12]) at a given sidereal time (in this case the trigger time of GW170817 [13, 14] event). The star represents the sky location of the GW170817 BNS event. We used PYCBC [15, 16] (a Python package to analyze CBC events) for generating these plots.

where $\overset{\leftrightarrow}{d}$ is detector tensor, defined as

$$D^{ab} = \frac{u^a u^b - v^a v^b}{2} \quad \text{or} \quad \frac{\vec{u} \otimes \vec{u} - \vec{v} \otimes \vec{v}}{2} \quad (1.76)$$

The detector tensor gives the response of the detector to a GW tensor $\overset{\leftrightarrow}{h}$. Using it, we can write down the strain measured by a detector as

$$\begin{aligned} h &= \overset{\leftrightarrow}{h} : \overset{\leftrightarrow}{D} \\ &= (h_+ \overset{\leftrightarrow}{e}^+ + h_\times \overset{\leftrightarrow}{e}^\times) : \overset{\leftrightarrow}{D} \\ &= h_+ F_+ + h_\times F_\times \end{aligned} \quad (1.77)$$

where F_+ and F_\times are the antenna pattern functions which are given by

$$F_+ = \overset{\leftrightarrow}{D} : \overset{\leftrightarrow}{e}^+ = D^{ab} e_{ab}^+ \quad (1.78a)$$

$$F_\times = \overset{\leftrightarrow}{D} : \overset{\leftrightarrow}{e}^\times = D^{ab} e_{ab}^\times \quad (1.78b)$$

For a given detector at a given time (i.e., for a fixed detector tensor $\overset{\leftrightarrow}{D}$), F_+ and F_\times depends on the three angles that define the sky location and polarization basis of the GW source with respect to some reference system. For example, in the equatorial coordinate system, they will depend on α , δ , and ψ . Given the sky location of the source, we can always construct a reference polarization basis $\{\overset{\leftrightarrow}{\varepsilon}^+, \overset{\leftrightarrow}{\varepsilon}^\times\}$ and then construct the following combinations:

$$a = \overset{\leftrightarrow}{D} : \overset{\leftrightarrow}{\varepsilon}^+ = D^{ab} \varepsilon_{ab}^+ \quad (1.79a)$$

$$b = \overset{\leftrightarrow}{D} : \overset{\leftrightarrow}{\varepsilon}^\times = D^{ab} \varepsilon_{ab}^\times \quad (1.79b)$$

using which we can calculate the antenna pattern functions:

$$F_+(\alpha, \delta, \psi) = a(\alpha, \delta) \cos 2\psi + b(\alpha, \delta) \sin 2\psi \quad (1.80a)$$

$$F_\times(\alpha, \delta, \psi) = -a(\alpha, \delta) \sin 2\psi + b(\alpha, \delta) \cos 2\psi \quad (1.80b)$$

From the observations in Fig. 1.3, it's clear that the antenna pattern functions exhibit a continuous variation across different sky locations. Furthermore, GW detectors have a wide sky coverage spanning $\sim 4\pi$, with only a few specific directions where they are blind. This is in contrast to conventional astronomy, where telescopes must be pointed very precisely towards a source for detection. However, it's important to note that a single GW detector cannot uniquely determine h_+ , h_\times , α , and δ . When using two detectors, we have access to two GW strains, $h_1(t)$ and $h_2(t)$, as well as a time delay τ_{12} . However, when employing three or more detectors, we can acquire at least three or more GW strains and two-time delays. With this additional information, we can confidently and uniquely determine all four unknowns.

1.1.4 GW generation in linearized theory

In the context of linearized theory, as discussed in Section 1.1.1, we assume that the source produces a sufficiently weak gravitational field. This allows us to perform an expansion around the flat space-time. This assumption further implies that within a system primarily influenced by gravitational forces, the typical velocities involved are relatively low. For instance, in a two-body system held together by gravitational interactions and characterized by a reduced mass μ and total mass M ,

$$\frac{1}{2}\mu v^2 = \frac{G\mu M}{r} \Rightarrow v^2/c^2 = R_S/2r \quad (1.81)$$

where $R_s = 2GM/c^2$ is the Schwarzschild radius corresponding to a mass M , from the above equation, we can deduce that in the presence of a weak gravitational field, the condition $R_s/r \ll 1$ holds, indicating that $v \ll c$. However, when dealing with systems primarily influenced by non-gravitational forces, we can explore weak field sources with velocities of arbitrary magnitudes. This simplifying assumption allows us to comprehend how gravitational waves (GWs) manifest in flat background space-time, which encompasses both Newtonian and special relativistic dynamics. Note that for many of the astrophysical sources of interest, such as neutron stars, black holes, and compact binary systems, which are inherently self-gravitating, the Newtonian approximation may not be applicable. In such cases, the Post-Newtonian (PN) [17, 18, 19] formalism becomes necessary for a comprehensive understanding of GW generation. However, we do not discuss the PN formalism here. Interested readers seeking a deeper exploration of this topic can refer to Michele Maggiore's book [7] and Luc Blanchet's Living Review in Relativity [20]. Now, let's turn to weak field sources characterized by arbitrary velocities while recalling the Eq. (1.18),

$$\square \bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4} T_{\mu\nu} \quad (1.82)$$

we found that in the Lorentz gauge, $\partial^\mu \bar{h}_\mu = 0$ and $T_{\mu\nu}$ satisfies that flat-space conservation law $\partial^\mu T_{\mu\nu} = 0$. Since the equation presented above is linear with respect to $h_{\mu\nu}$, we can employ

Green's function method (see Eq. (1.245) in Appendix 1.3) to find a solution. The solution to this equation in regions outside the source is as follows:

$$h_{ij}(t, \vec{x}) = \frac{4G}{c^4} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \int d^3x' \frac{1}{|\vec{x} - \vec{x}'|} T_{kl} \left(t - \frac{|\vec{x} - \vec{x}'|}{c}, \vec{x}' \right). \quad (1.83)$$

where $\Lambda_{ij,kl}$ is the Lambda tensor, which can be written in terms of Kronecker deltas as

$$\begin{aligned} \Lambda_{ij,kl}(\hat{\mathbf{n}}) &= \delta_{ik}\delta_{jl} - \frac{1}{2}\delta_{ij}\delta_{kl} - n_j n_l \delta_{ik} - n_i n_k \delta_{jl} \\ &+ \frac{1}{2}n_k n_l \delta_{ij} + \frac{1}{2}n_i n_j \delta_{kl} + \frac{1}{2}n_i n_j n_k n_l \end{aligned} \quad (1.84)$$

Here, we can expand $|\vec{x} - \vec{x}'|$ at $r \gg d$, where d is the typical radius of the source

$$|\vec{x} - \vec{x}'| = r - \vec{x}' \cdot \hat{\mathbf{n}} + \mathcal{O}\left(\frac{d^2}{r}\right), \quad (1.85)$$

To find the gravitational wave amplitude h_{ij} at the detector, which is far away from the source, we can take the limit $r \rightarrow \infty$ at fixed time t and keep only the leading order terms in the Eq. (1.83). Setting $|\vec{x} - \vec{x}'| = r$, at large distances, the Eq. (1.83) becomes

$$h_{ij}(t, \vec{x}) = \frac{4G}{rc^4} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \int d^3x' \frac{1}{|\vec{x} - \vec{x}'|} T_{kl} \left(t - \frac{r}{c} + \frac{\vec{x}' \cdot \hat{\mathbf{n}}}{c}, \vec{x}' \right), \quad (1.86)$$

and we write T_{kl} in terms of its Fourier transforms, we get

$$T_{kl} \left(t - \frac{r}{c} + \frac{\vec{x}' \cdot \hat{\mathbf{n}}}{c}, \vec{x}' \right) = \int \frac{d^4k}{(2\pi)^4} \tilde{T}_{kl}(\omega, \vec{k}) e^{-i\omega(t-r/c+\vec{x}' \cdot \hat{\mathbf{n}}/c)+i\vec{k} \cdot \vec{x}'}. \quad (1.87)$$

For non-relativistic sources, the Fourier transform of the stress-energy tensor, $\tilde{T}_{kl}(\omega, \vec{k})$ has a peak around a characteristic frequency ω_s , where $\omega_s d \ll c$. Additionally, inside the source, \tilde{T}_{kl} is non-vanishing, which means that the integral in the above equation is restricted to the region $\vec{x}' \leq d$. Therefore, the frequencies ω which contributes dominantly to h_{ij} satisfy

$$\frac{\omega}{c} \vec{x}' \cdot \hat{\mathbf{n}} \lesssim \frac{\omega_s d}{c} \ll 1, \quad (1.88)$$

Using the above condition in the Eq. (1.87), we get

$$T_{kl} \left(t - \frac{r}{c} + \frac{\vec{x}' \cdot \hat{\mathbf{n}}}{c}, \vec{x}' \right) \lesssim T_{kl} \left(t - \frac{r}{c}, \vec{x}' \right) + \frac{x'^i n^i}{c} \partial_0 T_{kl} + \frac{1}{2c^2} x'^i x'^j n^i n^j \partial_0^2 T_{kl} + \dots \quad (1.89)$$

where all the derivatives are evaluated at the point $(t - r/c, \vec{x}')$. Putting the above expansion in Eq. (1.86), we obtain

$$h_{ij}(t, \vec{x}) = \frac{4G}{rc^4} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \left[S^{kl} + \frac{1}{c} n_m \dot{S}^{kl,m} + \frac{1}{2c^2} n_m n_p \ddot{S}^{kl,mp} + \dots \right]_{\text{ret}}, \quad (1.90)$$

Here S represents various momentas of stress-energy tensor T^{ij} (see Eq. (1.246) in Appendix 1.3). The subscript “ret” indicates that all the quantities are evaluated at retarded time $t - r/c$. We also define second-order momentas of T^{00}/c^2 (which will be used later) as

$$M^{ij} = \frac{1}{c^2} \int d^x T^{00}(t, \vec{x}) x^i x^j, \quad (1.91)$$

It turns out that S^{ij} and M^{ij} are related by the following relation,

$$S^{ij} = \frac{1}{2} \ddot{M}^{ij}. \quad (1.92)$$

Now, using the above equation in Eq. (1.90) and keeping the leading term, we get

$$[h_{ij}(t, \vec{x})]_{\text{quad}} = \frac{2G}{rc^4} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \ddot{M}^{kl}(t - r/c). \quad (1.93)$$

where “quad” refers to the quadrupolar term. Denoting $\rho = T^{00}/c^2$ (in the lowest order in v/c , ρ becomes mass density) and defining the “quadrupole moment” as

$$\begin{aligned} Q^{ij} &\equiv M^{ij} - \frac{1}{3} \delta^{ij} M_{kk} \\ &= \int d^3x \rho(t, \vec{x}) (x^i x^j - \frac{1}{3} r^2 \delta^{ij}), \end{aligned} \quad (1.94)$$

and using the above equation in Eq. (1.93), we obtain

$$\begin{aligned} [h_{ij}(t, \vec{x})]_{\text{quad}} &= \frac{2G}{rc^4} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \ddot{Q}_{kl}(t - r/c) \\ &\equiv \frac{2G}{rc^4} \ddot{Q}_{ij}(t - r/c) \end{aligned} \quad (1.95)$$

For a case when the direction of propagation $\hat{\mathbf{n}}$ of GW to be $\hat{\mathbf{z}}$, the two polarization amplitudes can be written as

$$h_+ = \frac{G}{rc^4} (\ddot{M}_{11} - \ddot{M}_{22}) \quad (1.96a)$$

$$h_\times = \frac{G}{rc^4} \ddot{M}_{12}, \quad (1.96b)$$

It's important to highlight that the dominant term in Eq. (1.95) corresponds to the mass quadrupole and excludes monopole or dipole terms. To introduce the monopole and dipole terms, they would need to depend on mass density and momentum density, which are conserved quantities in linearized theory. Given that h_{ij} depends on the derivatives of the multipole moments, their contribution would vanish. Notably, the absence of these terms remains consistent even in the context of more generalized non-linear theory. Now, let's proceed to formulate the energy flux of GWs, which quantifies the energy carried by these waves per unit time through a unit surface located at a substantial distance from the source.

$$\frac{dE}{dt} = \frac{c^3 r^2}{32\pi G} \int d\Omega \langle \dot{h}_{ij} \dot{h}_{ij} \rangle \quad (1.97)$$

Substituting the Eq. (1.95) in the above equation, we find that in the quadrupole approximation, the radiated power per unit solid angle is given by

$$\begin{aligned} \left(\frac{dP}{d\Omega} \right)_{\text{quad}} &= \frac{r^2 c^3}{32\pi G} \langle \dot{h}_{ij} \dot{h}_{ij} \rangle \\ &= \frac{G}{8\pi c^5} \Lambda_{ij,kl}(\hat{\mathbf{n}}) \langle \ddot{Q}_{ij} \ddot{Q}_{ij} \rangle \end{aligned} \quad (1.98)$$

where the average is taken over several characteristic periods of the GW, and \ddot{Q}_{ij} is evaluated at the retarded time $t - r/c$. Using Eq. (1.247) in Appendix 1.3, we find the total radiative power

$$P_{\text{quad}} = \frac{G}{5c^5} \langle \ddot{Q}_{ij} \ddot{Q}_{ij} \rangle, \quad (1.99)$$

which is the Einstein's famous quadrupole formula [5]. Note that, in the quadrupole approximation, linear momentum is conserved. This can be seen by writing the rate of change of linear momentum as

$$\frac{dP^i}{dt} = -\frac{G}{8\pi c^5} \int d\Omega \ddot{Q}_{ab} \partial^i \ddot{Q}_{ab} \quad (1.100)$$

where the integrand is odd under reflection ($\vec{x} \rightarrow \vec{x}'$), owing to the fact that Q remains invariant under reflection while $\partial^i \rightarrow -\partial^i$. Consequently, the derivative dP^i/dt becomes zero. We conclude this section by showing quadrupole radiation from a mass in a circular orbit. Consider a binary system comprised of masses m_1 and m_2 where a circular motion is performed by the relative coordinate. Assuming the lowest order of v/c , which allows us to work in a flat space-time for describing a self-gravitating system. We choose a reference frame denoted as (x, y, z) , aligning with the binary orbit lying in the (x, y) plane, which is described by the following equation:

$$\begin{aligned} x_0(t) &= R \cos(\omega_s t + \pi/2) \\ y_0(t) &= R \sin(\omega_s t + \pi/2) \\ z_0(t) &= 0 \end{aligned} \quad (1.101)$$

The second mass moment for this binary system is $M^{ij} = \mu x_0^i(t)x_0^j(t)$, where $\mu = m_1m_2/(m_1+m_2)$ is the reduced mass of the system, so the non-vanishing components are

$$\begin{aligned} M_{11} &= \mu R^2 \frac{1 - \cos 2\omega_s t}{2}, \\ M_{22} &= \mu R^2 \frac{1 + \cos 2\omega_s t}{2}, \\ M_{12} &= -\frac{1}{2} \mu R^2 \sin(2\omega_s t), \end{aligned} \quad (1.102)$$

Using the above equations, we get $\ddot{M}_{11} = 2\mu R^2 \omega_s^2 \cos(2\omega_s t)$ and $\ddot{M}_{12} = 2\mu R^2 \omega_s^2 \sin(2\omega_s t)$ while $\ddot{M}_{22} = -\ddot{M}_{11}$. Using these expressions (see Eq. (3.72) in [7]), we find

$$h_+(t; \theta, \phi) = \frac{4G\mu\omega_s^2 R^2}{rc^4} \left(\frac{1 + \cos^2 \theta}{2} \right) \cos(2\omega_s t_{\text{ret}} + 2\phi), \quad (1.103a)$$

$$h_\times(t; \theta, \phi) = \frac{4G\mu\omega_s^2 R^2}{rc^4} \cos \theta \sin(2\omega_s t_{\text{ret}} + 2\phi) \quad (1.103b)$$

Notice that the quadrupole radiation occurs at a frequency that is precisely "twice" the source frequency ω_s . When observed, we only detect the radiation that is emitted in our direction from the binary system. Therefore, the angle θ corresponds to the inclination angle ι between the normal to the orbital plane and our line of sight. To simplify the equations, we can fix ϕ at zero. Consequently, the equations above can be expressed as follows (see [7] for detail):

$$h_+(t; \theta, \phi) = \frac{4G\mu\omega_s^2 R^2}{rc^4} \left(\frac{1 + \cos^2 \iota}{2} \right) \cos(2\omega_s t_{\text{ret}}), \quad (1.104a)$$

$$h_\times(t; \theta, \phi) = \frac{4G\mu\omega_s^2 R^2}{rc^4} \cos \iota \sin(2\omega_s t_{\text{ret}}) \quad (1.104b)$$

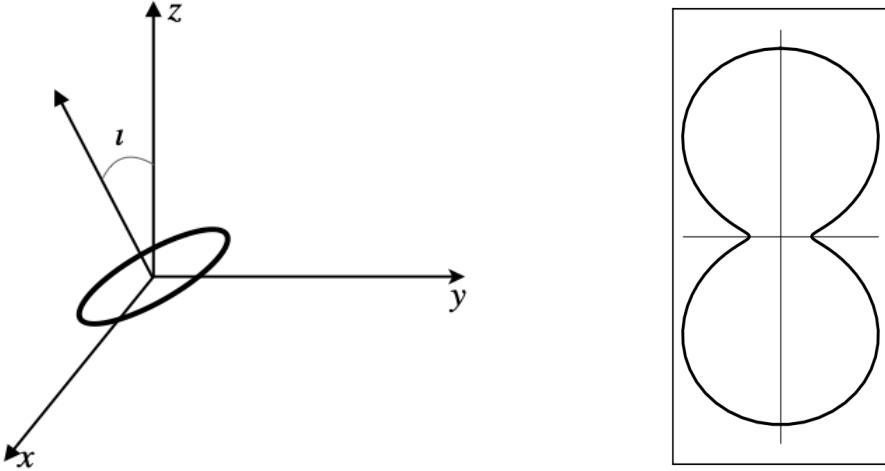


Figure 1.4: The left panel shows the coordinate system chosen for the binary system we have considered here. The right panel shows $g(\iota)$ in polar coordinates where ι is measured from the vertical axis (reproduced from [7]).

For $\iota = 0$ (corresponding to a face-on orbit), both h_+ and h_\times have identical amplitude, and GW becomes circularly polarized. In contrast, when ι equals $\pi/2$ (representing an edge-on orbit), h_\times becomes zero, and the gravitational wave radiation becomes linearly polarized. For $0 \leq \iota \leq \pi/2$, the h_+ and h_\times amplitudes are different, causing the gravitational waves to exhibit elliptical polarization. It's worth noting that by measuring the relative amplitudes of these two polarizations, the inclination angle ι of the orbit can be estimated. Using Eq. (1.98) and insert Eq. (1.104), the angular distribution of the radiative power is

$$\left(\frac{dP}{d\Omega}\right)_{\text{quad}} = \frac{2G\mu^2 R^4 \omega_s^6}{\pi c^5} g(\iota), \quad (1.105)$$

where $g(\iota) = (1 + \cos^2 \iota)^2 / 2 + \cos^2 \iota$. It is worth mentioning that $g(\theta)$ never vanishes irrespective of the inclination angle ι , implying the non-vanishing component of the source's motion perpendicular to the light of sight of the observer as evident from Fig. 1.4. In the next section, we will discuss the inspiral of compact binaries, which are the most interesting astrophysical sources of GWs.

1.1.5 Inspiring compact binaries

In the preceding section, we considered a scenario involving a binary system comprising two point masses, denoted as m_1 and m_2 , which were in a circular orbit. By invoking Kepler's law, we can establish a relation between the orbital frequency ω_s and the orbital radius R as follows:

$$\omega_s^2 = \frac{GM}{R^3} \quad \Rightarrow \quad R = \left(\frac{GM}{\omega_s^2}\right)^{1/3} \quad (1.106)$$

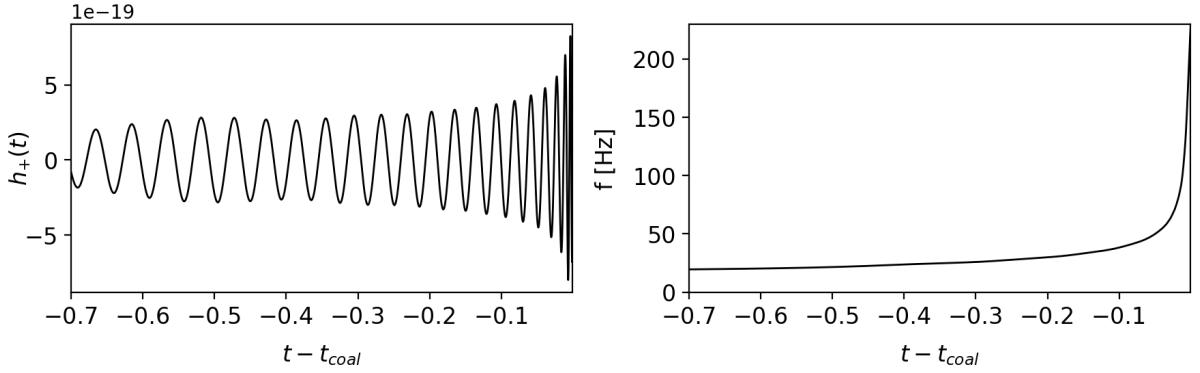


Figure 1.5: *Left panel:* The time evolution of the GW amplitude $h_+(t)$ in the inspiral phase of compact binary of equal mass of $35 M_\odot$. The binary spends ~ 0.7 seconds in the frequency band starting from 20 Hz before merging. *Right panel:* The corresponding frequency evolution of the GW signal exhibiting a typical chirp-like behavior. We used PYCBC [15, 16] (a Python package to analyze CBC events) for generating these plots.

We substitute R from the above equation in Eq.(1.103) and write $\omega_{\text{gw}} = 2\omega_s = 2\pi f_{\text{gw}}$ and find

$$h_+(t) = \frac{4}{r} \left(\frac{G\mathcal{M}}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}}{c} \right)^{2/3} \frac{1 + \cos^2 \iota}{2} \cos(2\pi f_{\text{gw}t_{\text{ret}}} + 2\phi) \quad (1.107\text{a})$$

$$h_\times(t) = \frac{4}{r} \left(\frac{G\mathcal{M}}{c^2} \right)^{5/3} \left(\frac{\pi f_{\text{gw}}}{c} \right)^{2/3} \cos \iota \sin(2\pi f_{\text{gw}t_{\text{ret}}} + 2\phi), \quad (1.107\text{b})$$

where \mathcal{M} is the chirp mass $\mathcal{M} = \mu^{3/5} M^{2/5} = (m_1 m_2)^{3/5} / (m_1 + m_2)^{1/5}$. Note that in the quadrupole approximation, the amplitudes h_+ and h_\times (see Fig. 1.5) of the GWs emitted depend on the masses m_1 and m_2 only through the combination \mathcal{M} . Similarly, we write the Eq. (1.105) in terms of \mathcal{M} and ω_{gw}

$$\frac{dP}{d\Omega} = \frac{2}{\pi} \frac{c^5}{G} \left(\frac{G\mathcal{M}\omega_{\text{gw}}}{2c^3} \right)^{10/3} g(\iota), \quad (1.108)$$

The total radiated power P is therefore

$$P = \frac{32}{5} \frac{c^5}{G} \left(\frac{G\mathcal{M}\omega_{\text{gw}}}{2c^3} \right)^{10/3}. \quad (1.109)$$

where we have used the fact that the angular average of $g(\iota)$ is $4/5$. As the GWs are emitted, the orbital energy decreases, implying the R must decrease over time. This reduction in R results in an increase in ω_s , which leads to a further increase in the radiated power. It starts a runaway process which, on a sufficiently long time scale, results in a coalescence of the binary system. As long as $\dot{\omega}_s \ll \omega_s^2$, the binary is in the regime called “quasi-circular” which can be translated to $\dot{R} = -2/3(\omega_s R)\dot{\omega}_s/\omega_s^2$ where $|\dot{R}| \ll \omega_s R$ as long as the quasi-circular condition holds. To calculate the rate of change of the GW frequency, we can equate the rate of change of orbit energy \dot{E}_{orbit} to the total power radiated given in the Eq. (1.109)

$$\dot{f}_{\text{gw}} = \frac{96}{5} \pi^{8/3} \left(\frac{G\mathcal{M}}{c^3} \right)^{5/3} f_{\text{gw}}^{11/3}. \quad (1.110)$$

which, when written in terms of the time to coalescence $\tau \equiv t_{\text{coal}} - t$ with t as the observer time and t_{coal} as the time of coalescence, is given by

$$f_{\text{gw}}(\tau) = \frac{1}{\pi} \left(\frac{5}{256 \tau} \right)^{3/8} \left(\frac{G\mathcal{M}}{c^3} \right)^{-5/8}. \quad (1.111)$$

where f_{gw} diverges at the time of coalescence f_{gw} beyond which the quadrupole approximation breaks down. Using the above equation, we can calculate the time to coalescence τ

$$\tau \simeq 2.18s \left(\frac{1.21 M_{\odot}}{\mathcal{M}} \right)^{5/3} \left(\frac{100 \text{ Hz}}{f_{\text{gw}}} \right)^{8/3}. \quad (1.112)$$

To estimate the typical time to coalescence, τ , we consider a binary system of two neutron stars with equal masses $m_1 = m_2 = 1.4 M_{\odot}$ at the $f_{\text{gw}} = 10 \text{ Hz}$, we get around $\tau = 17 \text{ min}$ to coalescence. We can also calculate the number of cycles \mathcal{N}_{cyc} spent within a frequency band accessible to a detector as

$$\mathcal{N}_{\text{cyc}} = \int_{f_{\text{min}}}^{f_{\text{max}}} df_{\text{gw}} \frac{f_{\text{gw}}}{\dot{f}_{\text{gw}}} \quad (1.113a)$$

$$\simeq 1.6 \times 10^4 \left(\frac{10 \text{ Hz}}{f_{\text{min}}} \right)^{5/3} \left(\frac{1.2 M_{\odot}}{\mathcal{M}} \right)^{5/3}, \quad (1.113b)$$

From the equation above, we can infer that ground-based GW detectors can access thousands of cycles throughout the GW signal's evolution, whereas space-borne detectors can record millions of cycles, highlighting the importance of precisely modeling GW waveforms. When considering the GW's evolution, we can reasonably assume that \dot{R} is negligible, as long as $f_{\text{gw}} \ll 13 \text{ kHz} (1.2 M_{\odot}/\mathcal{M})$. Consequently, we can reframe Eq. (1.107) in terms of the observer's measurement of the time to coalescence τ .

$$h_+(t) = \frac{1}{r} \left(\frac{G\mathcal{M}}{c^2} \right)^{5/4} \left(\frac{5}{c\tau} \right)^{1/4} \frac{1 + \cos^2 \iota}{2} \cos[\Phi(\tau)] = A_+ \cos[\Phi(\tau)] \quad (1.114a)$$

$$h_{\times}(t) = \frac{1}{r} \left(\frac{G\mathcal{M}}{c^2} \right)^{5/4} \left(\frac{5}{c\tau} \right)^{1/4} \cos \iota \sin[\Phi(\tau)] = A_{\times} \sin[\Phi(\tau)], \quad (1.114b)$$

where $\Phi(\tau) = -2 \left(\frac{5G\mathcal{M}}{c^3} \right)^{-5/8} \tau^{5/8} + \Phi_0$, and Φ_0 is the phase at coalescence. As evident from the above equations, it is clear that both the frequency and amplitude of the signal increase as the binary system approaches its merger. Up to this point, all calculations have been conducted under the assumption of a flat background. In a realistic scenario involving compact binaries like binary neutron stars or binary black holes, the gravitational field is significantly influenced by the proximity of these massive objects, impacting the binary's dynamics. In accordance with the Schwarzschild geometry, there exist a minimum radial distance beyond which a stable circular orbit is not possible. This critical boundary is known as the Innermost Stable Circular Orbit (ISCO) and is defined as:

$$r_{\text{ISCO}} = \frac{6GM}{c^2} \quad (1.115)$$

where M is the total mass of the binary. The corresponding frequency beyond which the condition of “quasi-circular” orbits no longer holds can be written as

$$f_{\text{ISCO}} = \frac{1}{6\sqrt{6}(2\pi)} \frac{c^3}{GM} \simeq 2.2 \text{kHz} \left(\frac{M_\odot}{M} \right) \quad (1.116)$$

For a binary neutron star system with both masses being approximately $1.4 M_\odot$, the ISCO frequency, denoted as f_{ISCO} , is estimated to be ~ 800 Hz. In the case of a binary black hole system where both masses are roughly $15 M_\odot$, the ISCO frequency is ~ 73 Hz. In conclusion, we will now write down the Fourier transform of the chirp signal presented in Eq. (1.114). The corresponding Fourier transforms are as follows (see [7] for derivation):

$$\tilde{h}_+(f) = Ae^{i\Psi_+(f)} \frac{c}{r} \left(\frac{G\mathcal{M}}{c^3} \right)^{5/6} \frac{1}{f^{7/6}} \left(\frac{1 + \cos^2 \iota}{2} \right), \quad (1.117a)$$

$$\tilde{h}_\times(f) = Ae^{i\Psi_\times(f)} \frac{c}{r} \left(\frac{G\mathcal{M}}{c^3} \right)^{5/6} \frac{1}{f^{7/6}} \cos \iota, \quad (1.117b)$$

where $A = 1/\pi^{2/3}(5/24)^{1/2}$ and phases are given by

$$\Psi_+(f) = 2\pi f(t_c + r/c) - \Phi_0 - \frac{\pi}{4} + \frac{3}{4} \left(\frac{8\pi f G \mathcal{M}}{c^3} \right)^{-5/3}, \quad (1.118)$$

where Φ_0 is phase of coalescence, with phase for cross-polarization is $\Psi_\times = \Psi_+ + \pi/2$. The next section will discuss the first indirect and direct evidence of GWs.

1.1.6 Indirect evidence and direct evidence of GWs

Indirect evidence: Hulse-Taylor binary pulsar

In July 1974, Hulse and Taylor made a breakthrough discovery when they detected a pulsar called PSR B1913+16, also known as the Hulse-Taylor binary pulsar [21, 22, 23]. This pulsar, characterized by a period of approximately 59 ms, was observed using the Arecibo Observatory in Puerto Rico. Pulsars are rapidly rotating neutron stars with intense magnetic fields that emit electromagnetic radiation in the form of a beam through their magnetic poles. This radiation becomes observable to us only when the beam is oriented toward Earth, much like a lighthouse effect. The Hulse-Taylor binary pulsar exhibits an orbital period of around 7.6 hours and an orbital velocity of approximately 10^{-3} times the speed of light (c), indicating its relativistic nature. An intriguing aspect of this observation is the rate of periastron advance, denoted as $\dot{\omega}$ which is ~ 4.22 degrees per year, significantly larger than the perihelion advance of Mercury, which is ~ 43 arcseconds per century [24, 25]. This highlights the substantial impact of GR effects in the Hulse-Taylor pulsar system. Through the determination of $\langle \dot{\omega} \rangle$, where ω represents the angular position of the periastron, measured from the ascending node and the Einstein parameter γ , which characterizes second-order Doppler-gravitational-redshift timing, the masses of the pulsar (m_p) and its companion (m_c) can be measured and found to be approximately $1.4412 M_\odot$ and $1.3867 M_\odot$, respectively. Utilizing these values of masses along with the orbital

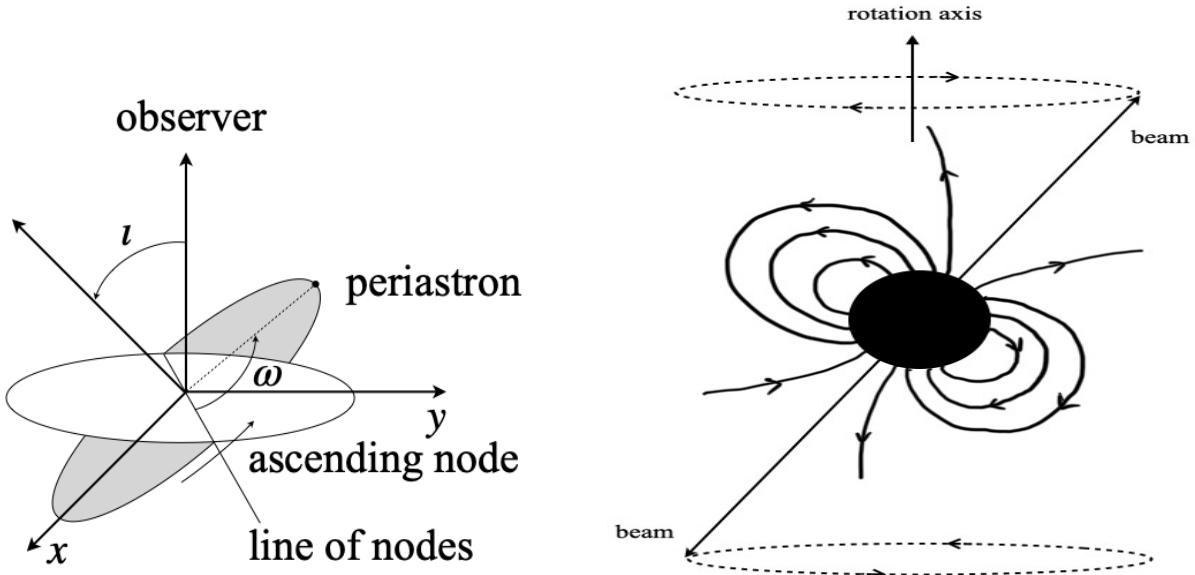


Figure 1.6: *Left panel:* The pulsar’s orbit geometry where the orbital plane is shown in grey. *Right panel:* A schematic of the magnetosphere of the pulsar and the beams of outgoing radiation (reproduced from [7]).

period, the semimajor axis (a) is calculated using Kepler’s formula, yielding an estimate of roughly $3 R_{\odot}$. The compactness of the orbit and the absence of eclipses during observations strongly suggest that the companion is a compact star, possibly a binary black hole (BBH) or binary neutron star (BNS). However, the value of m_c indicates that it is more likely a BNS. Finally, \dot{P} , the rate of change of the orbital period, can be measured by combining information about the masses and eccentricity, using the predictions of GR regarding the decrease in the orbital period due to the emission of GW. The ratio of the experimentally determined value of \dot{P}_b ($(\dot{P}_b)_{\text{exp}}$) to the value predicted by GR ($(\dot{P}_b)_{\text{GR}}$) comes out to be

$$\frac{(\dot{P}_b)_{\text{exp}}}{(\dot{P}_b)_{\text{GR}}} \sim 1.0013 \quad (1.119)$$

This discovery provided a remarkable validation of GR and the existence of GWs. Next, we will discuss the process of detecting the signals emitted by the pulsar and explore the various factors influencing the detection.

Pulsar as stable clocks As previously mentioned, pulsars are rapidly rotating, magnetized neutron stars with short rotational periods on the order of milliseconds, a characteristic resulting from the conservation of angular momentum. Additionally, the conservation of magnetic flux leads to the presence of extremely strong magnetic fields, reaching magnitudes as high as 10^{12} gauss. These powerful magnetic fields influence the behavior of charged particles in the vicinity of the pulsar. Within a critical distance from the rotation axis, denoted as $\rho_c = c/\Omega$ (where ρ represents the distance from the rotation axis in cylindrical coordinates), magnetic field lines are closed. Beyond this critical radius, at $\rho > \rho_c$, magnetic field lines are open, allowing charged particles to escape into space. This region, known as the magnetosphere, encompasses a cylindrical volume

with a radius of ρ_c and is filled with ionized, high-energy plasma that co-rotates with the neutron star. In the magnetosphere, the high-energy particles follow a constrained path along the open magnetic field lines. This motion gives rise to EM radiation, which is primarily observed in the form of narrow radio waves emitted from the magnetic poles of the pulsar. Each of these radiation beams traces out a circle in the sky, and if one of these beams intersects the observer's line of sight on Earth, it results in the detection of pulses. To study and characterize pulsars, a coherent addition of these pulses is performed, creating a single, stable profile that acts as a unique fingerprint for each pulsar. The resulting averaged profile becomes a template for further analysis. A least square fit is employed to obtain the time-of-arrival (TOAs) of the individual pulses. The data consisting of TOAs, collected over time, offers remarkable precision and allows researchers to determine the exact number of pulses that occurred between the closure of the Arecibo telescope in the mid-1990s and the present day.

While the times of arrival (TOAs) of these pulses from pulsars demonstrate remarkable stability, their precision can be influenced by various time-dependent factors. These factors include the Earth's motion around the Sun and the gravitational effects of the solar system as predicted by general relativity. Additionally, in cases where the pulsar is part of a binary system, there are further modulations to consider when calculating the TOAs of the pulses. Since TOAs are dependent on the specific parameters of the binary system, such as the component masses, their measurements can be used to estimate these parameters. Now, we will briefly discuss various factors affecting the TOAs of the pulses.

Time-delays and Dispersion There are three kinds of delays that can affect the TOAs of the pulses arriving from the pulsars: (i) Roemer time delay, (ii) Shapiro time delay, and (iii) Einstein time delay. The Roemer time delay is given by

$$\Delta_{R,\odot} = -\vec{r}_{\text{ob}} \cdot \hat{\mathbf{s}}/c, \quad (1.120)$$

where $r_{\text{ob}} = r_{\text{oe}} + r_{\text{es}} + r_{\text{sb}}$. Here r_{oe} is the vector pointing from the observer to the Earth's center, r_{es} from the Earth's center to the Sun's center, r_{sb} from the Sun's center to the solar system barycenter (SSB), and $\hat{\mathbf{s}}$ is the unit vector from the SSB to the pulsar. Taking into account that light's propagation is also affected by the general relativistic effects due to the solar system's gravitational field, the arrival time t_{obs} at the observer is given by

$$t_{\text{SSB}} = t_{\text{obs}} + \Delta_{R,\odot} + \frac{2}{c} \int_{\vec{r}_{\text{obs}}}^{\vec{r}_p} |d\vec{x}| \phi(\vec{x}). \quad (1.121)$$

In this equation, $\phi(\vec{x})$ represents the metric perturbation, and $t_{\text{SSB}} = t_e + |\vec{r}_p - \vec{r}_b|/c$ corresponds to the time when the pulse would reach the Solar System Barycenter (SSB) without the influence of the solar system's gravitational effects. Here, \vec{r}_p and \vec{r}_b denote the positions of the pulsar and the SSB, respectively. The final term, preceded by a negative sign, denotes the Shapiro time delay [26] due to the solar system, represented as $\Delta_{S,\odot}$. It's important to note that a laboratory clock measures its proper time τ rather than the coordinate time t utilized in the calculations of the previous two delays. Using the relationship between coordinate time t and proper time τ , we

can express this as:

$$t \simeq \tau + \int^t dt' \left[\frac{v_{\text{obs}}^2(t')}{2c^2} - \phi(\vec{x}_{\text{obs}}(t')) \right] \quad (1.122)$$

where second term represents the Einstein time delay $\Delta_{E,\odot}$. In addition to the previously mentioned delays, there is another correction applied to the TOAs due to the radio waves traveling through ionized interstellar gas, causing the waves to disperse. This dispersion effect correction is expressed as $-D/\nu^2$, with $D = (e^2/2\pi m_e c)$ representing the Dispersion Measure (DM) and ν indicating the frequency of the radio pulse. The Dispersion Measure, often referred to as DM, is calculated as $\text{DM} \equiv \int_0^L n_e dl$, where n_e signifies the electron number density, and L is the distance traveled. To account for all these factors, we can combine the various delays and dispersion effects as follows:

$$t_{\text{SSB}} = \tau_{\text{obs}} - \frac{D}{\nu^2} + \Delta_{E,\odot} + \Delta_{R,\odot} - \Delta_{S,\odot}. \quad (1.123)$$

where t_{SSB} is the coordinate time at which laboratory clock records the signal, and τ_{obs} is observer's proper time. Up to this point, we have exclusively focused on the case of a single pulsar emitting electromagnetic pulses. In scenarios where the pulsar is part of a binary system, there exist additional delays and corrections that must be integrated into the final timing formula. However, the detailed treatment of such complexities falls outside the scope of this thesis. Readers interested in exploring these intricacies further are encouraged to consult Section 6.2.3 of Michele Maggiore's Volume 1 [7]. To conclude this section, we highlight several crucial observations stemming from the Hulse-Taylor binary system. Through a least-squares fitting of Time of Arrivals (TOAs), we can extract all the required parameters of the binary system, accounting for delays, dispersion, and necessary corrections. Referring to Equation (1.119), we identify a remarkable alignment between the observed rate of change in the orbital period and the value predicted by General Relativity. The corresponding accumulated phase ϕ_b is

$$\frac{\phi_b(T)}{2\pi} = \nu_b T + \frac{1}{2} \dot{\nu}_b T^2 + \dots, \quad (1.124)$$

where ν_b is the orbital frequency, and T is the proper time in the pulsar frame. The n -th time of periastron passage, T_n is defined by the condition, $\phi_b(T_n) = 2\pi n$

$$\nu_b T_n + \frac{1}{2} \dot{\nu}_b T_n^2 = n. \quad (1.125)$$

The cumulative difference between the periastron passages T_n and the values nP_b can be written as follows:

$$T_n - nP_b = \frac{\dot{P}_b}{2P_b} T_n^2. \quad (1.126)$$

which is equation of a parabola with $\dot{P}_b/(2P_b) < 0$ as the coefficient. As depicted in Fig. 1.7, the agreement between the theoretical predictions and the experimental data is exceptional and has demonstrated remarkable consistency over the years. Recent assessments of the Hulse-Taylor pulsar suggest that it will cease to be observable after the year 2025 [27, 28]. This is due to its geodetic precession, which arises from the effects of general relativity associated with the spin-orbit coupling.

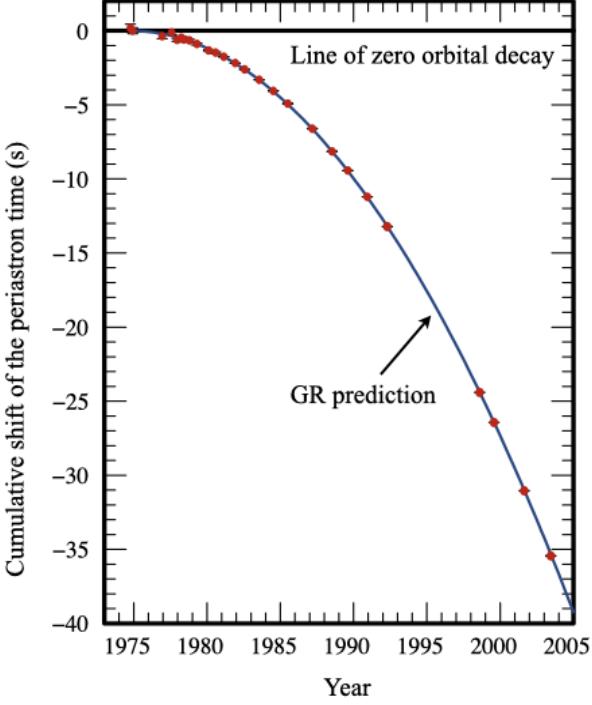


Figure 1.7: The plot depicts the orbital decay of the Hulse-Taylor pulsar PSRB1913+16 [21, 29]. The red dots represent the observed variations in the epoch of periastron with time. Meanwhile, the parabolic curve, illustrated by the solid blue line, represents the GR prediction for a system emitting gravitational radiation.

Direct evidence of GWs

GW150914 BBH event On September 14, 2015, approximately a century after Einstein’s initial prediction, a significant milestone was achieved when the two ground-based LIGO detectors, LIGO-Livingston and LIGO-Hanford [12], jointly observed a simultaneous gravitational wave event referred to as “GW150914” [30, 2]. This event originated from a binary black hole system with source component masses of $\sim 36 M_{\odot}$ for one black hole and $\sim 29 : M_{\odot}$ for the other, resulting in a final black hole with a mass of approximately $62 M_{\odot}$ and a final black hole spin of $a_f \sim 0.67$. The source was estimated to be situated at a distance of approximately 410 million parsecs, corresponding to a redshift of around $z \sim 0.099$. The detection of GW150914 marked a historic moment, offering the first direct confirmation of GW’s existence and reaffirming the validity of Einstein’s theory of General Relativity. This breakthrough ushered in a new era of gravitational wave astronomy.

GW170817 BNS event: First multimessenger event The GW170817 binary neutron star (BNS) event marked a significant milestone as the first multimessenger event detected by LIGO detectors [12, 31], during the second observation run (O2) [13]. Its electromagnetic (EM) counterpart was also observed by various EM facilities worldwide [32, 33]. This event had a combined Signal-to-Noise Ratio (SNR) of 32.4, featuring component masses of approximately $1.4 M_{\odot}$ and $1.2 M_{\odot}$ and component spins that were nearly zero. It originated from a host

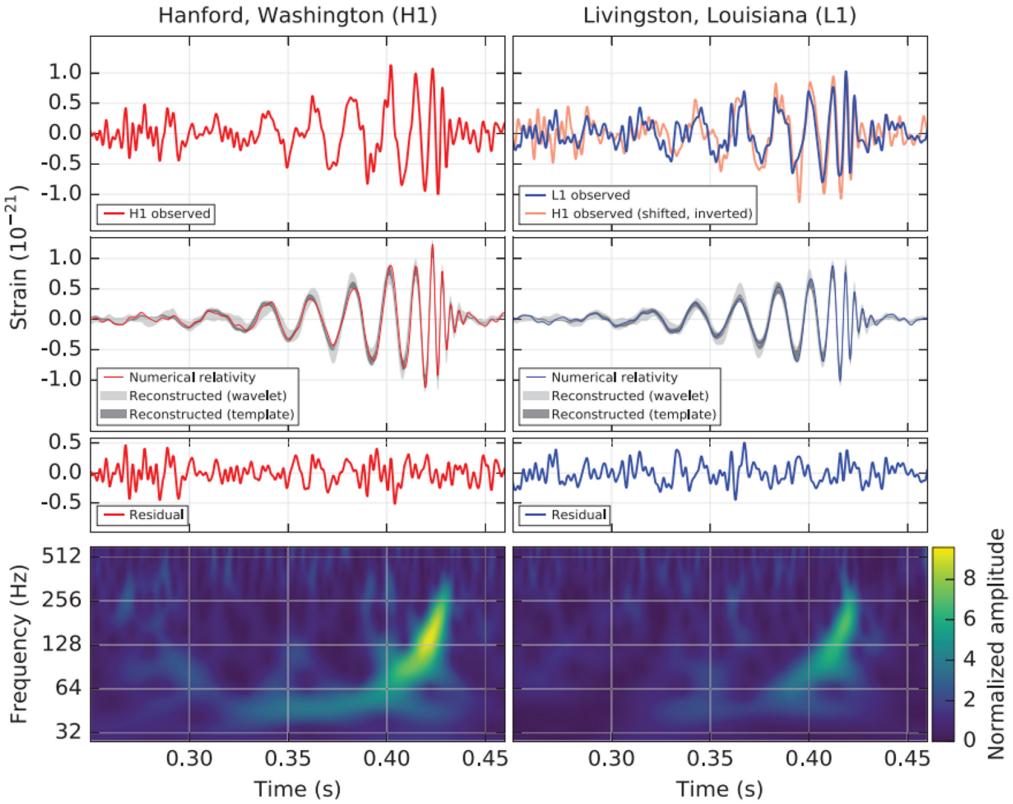


Figure 1.8: *Top row:* The GW strain (bandpassed) as observed by H1 (left) and L1 (right). *Second row:* Gravitational-wave strain projected onto each detector in the 35–350 Hz band. The solid red line represents the numerical relativity waveform generated for a system with parameters consistent with those recovered from GW150914. Dark (light) grey areas show 90% credible region of the signal reconstructed using BBH template waveforms (a linear combination of sine-Gaussian wavelets). *Third row:* Residuals after subtracting the filtered numerical relativity waveform from the filtered detector time series. *Bottom row:* A time-frequency map of the strain data, showing a typical chirp behavior of the BBH signal (taken from [30]).

galaxy known as “NGC4993” [34] positioned at a distance of roughly 40 million parsecs from Earth. Approximately 1.7 seconds after the GW detection, a gamma-ray burst was also observed, which played a crucial role in constraining the propagation speed of gravitational waves. These multimessenger events provide valuable insights into constraining the physical models governing the emission processes. Consequently, early detection of such events becomes imperative, necessitating the development of an accurate and rapid parameter estimation technique to estimate the source’s sky location and component masses. One such technique will be elaborated upon in Chapter 4, where we discuss the rapid parameter estimation for BNS/NSBH sources.

1.1.7 Detection strategies

In the preceding sections, we examined the generation and sources of GWs. In this section, we will explore methods for detecting these signals. Given that GW signals are exceedingly

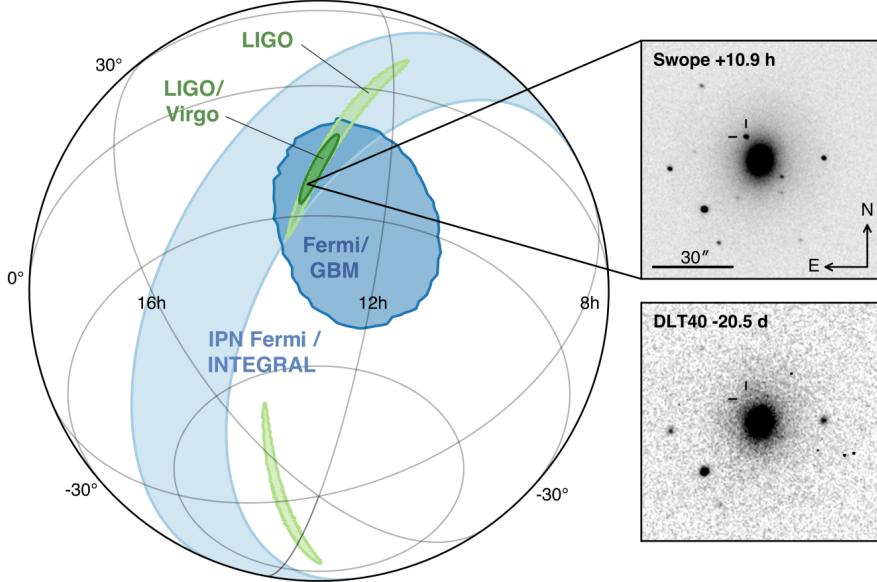


Figure 1.9: The localization of the GW170817 event, along with the gamma-ray and optical signals, is illustrated in the figure. In the left panel, skymaps indicating the 90% credible regions are displayed. These regions are represented by light green for LIGO (dark green for LIGO-Virgo), IPN triangulation based on the time delay between Fermi and INTEGRAL (light blue), and Fermi-GBM. Additionally (dark blue), the inset provides a view of the apparent location of the host galaxy NGC 4993. It is also depicted using the Swope optical discovery image captured approximately 10.9 hours after the merger (top right). The bottom right image shows the DLT40 per-discovery image taken roughly 20.5 days prior to the merger (taken from [32]).

faint and often obscured by various types of significantly stronger noise sources (as discussed in Section 1.1.8), it is imperative to develop optimal techniques to extract these signals from the noise. The output of a GW detector typically takes the form of a time series that describes various parameters, such as the motion of test masses or phase shifts in light. Naturally, we anticipate that this output contains the real signal along with the noise. A GW detector can be viewed as a linear system, where the input represents the GW signal $h(t) = D^{ij}h_{ij}$, with h_{ij} characterizing the GW and D^{ij} denoting the detector's sensor (as detailed in Section 1.1.3), which is dependent upon the detector's geometry. The resulting output is a time series expressed as follows:

$$d(t) = h(t) + n(t) \quad (1.127)$$

Assuming the noise is stationary and the corresponding Fourier components uncorrelated, their ensemble average is given by

$$\langle \tilde{n}^*(f)\tilde{n}(f') \rangle = \delta(f - f') \frac{1}{2}S_n(f) \quad (1.128)$$

where $S_n(f)$ is called the “single-sided” power spectral density and it has dimensions Hz^{-1} . The term single-sided comes from the fact that only positive frequencies are used to calculate the

$\langle n^2(t) \rangle$. From the above equation, we can write

$$\langle |\tilde{n}(f)|^2 \rangle = \frac{1}{2} S_n(f) T \quad (1.129)$$

where we are restricting ourselves to a time interval $-T/2 < t < T/2$ for calculating $\tilde{n}(f)$ which implies that the frequency resolution is given by $\Delta f = 1/T$. Now, we will discuss the technique of matched filtering in the context of GWs.

Matched filtering

As discussed in the previous section, we can express the detector's output as a combination of gravitational wave strain, denoted as $h(t)$, and noise, represented as $n(t)$. Typically, the noise component exhibits significantly greater magnitude compared to $h(t)$, especially considering Earth-based detections. This raises a crucial question: how can we detect such faint signals that are deeply embedded within the noise? Surprisingly, we can indeed detect these signals without requiring that $|h(t)| \ll |n(t)|$. It is sufficient for h_0 to satisfy the condition $h_0 > (\tau_0/T)^{1/2} n_0$, where h_0 and n_0 are the characteristic amplitudes of $h(t)$ and $n(t)$, τ_0 is the typical characteristic time of $h(t)$, and T is the duration of the observation. Defining

$$\hat{d} = \int_{-\infty}^{\infty} dt d(t) K(t), \quad (1.130)$$

where $K(t)$ is the filter function. Now, we need to estimate optimal $F(t)$, which maximizes the signal-to-noise ratio (SNR), denoted as ρ_m . It is the ratio of the expected value of \hat{d} in the presence of a signal and the expected value of the \hat{d} when the signal is not present. Assuming $\langle n(t) \rangle = 0$, the ρ_m is given by,

$$\rho_m = \frac{\int_{-\infty}^{\infty} df \langle \tilde{h}(f) \rangle \tilde{K}^*(f)}{\left[\int_{-\infty}^{\infty} df (1/2) S_n(f) |\tilde{K}(f)|^2 \right]^{1/2}} \quad (1.131)$$

The above equation can be succinctly written as

$$\rho_m = \frac{\langle u | h \rangle}{\langle u | u \rangle^{1/2}}. \quad (1.132)$$

where $u(t)$ is a function whose Fourier transform is given by $\tilde{u}(f) = S_n(f) \tilde{K}(f)$ and $\langle \cdot | \cdot \rangle$ denotes the inner product ². It is clear that u , when chosen parallel to the h , gives the maximum value of ρ_m . It implies that the optimal filter (also known as the matched filter [35, 36]) is given by

$$\tilde{K}(f) \propto \frac{\tilde{h}(f)}{S_n(f)} \quad (1.133)$$

Using the above equations, we find the optimal SNR denoted as ρ_{opt} as

$$\rho_{\text{opt}} = \sqrt{4 \int_0^{\infty} df \frac{|\tilde{h}(f)|^2}{S_n(f)}}, \quad (1.134)$$

²The inner product is defined as $\langle x | y \rangle = 4 \operatorname{Re} \int_0^{\infty} df \frac{\tilde{x}^*(f) \tilde{y}(f)}{S_n(f)}$

Source reconstruction

Maximum likelihood estimate There are primarily two approaches to source reconstruction: (i) the frequentist approach and (ii) the Bayesian approach. In the context of parameter estimation for GW sources, we adopt the Bayesian approach because we lack an ensemble of identical GW events, unlike particle physics experiments where experimenters can manipulate various experiment parameters. The Bayesian approach is advantageous as it can provide answers to questions such as the most probable values of the source parameters given the GW data. Now, let's consider a scenario in which a GW detection has been made. This implies that for a particular template $h(t, \vec{\theta})$, where $\vec{\theta}$ represents various source parameters like the masses and spins of the binary, etc., the signal-to-noise ratio, denoted as ρ_m , has surpassed a predetermined threshold. At this point, we aim to estimate the source parameters. We can achieve this by estimating the posterior distribution over the source parameters $\vec{\theta}$. Utilizing Eq. (1.128), we can express the probability distribution of noise as:

$$\begin{aligned} P(n_0) &= \mathcal{N} \exp \left[-\frac{1}{2} \int_{-\infty}^{\infty} df \frac{|\tilde{n}_0(f)|^2}{(1/2)S_n(f)} \right] \\ &= \mathcal{N} \exp [-\langle n_0 | n_0 \rangle / 2]. \end{aligned} \quad (1.135)$$

where we assumed a stationary and Gaussian noise $n(t)$ and \mathcal{N} is the normalization constant. Since $d(t) = h(t, \vec{\theta}_t) + n_0(t)$, plugging $n_0(t)$ back in the above equation, we get

$$\begin{aligned} P(d | \theta_t) &= \mathcal{N} \exp \left[-\frac{1}{2} \langle d - h(\theta_t) \rangle | \langle d - h(\theta_t) \rangle \right], \\ &= \mathcal{N} \exp \left[\langle h(\theta_t) | d \rangle - \frac{1}{2} \langle h(\theta_t) | h(\theta_t) \rangle - \langle d | d \rangle \right] \end{aligned} \quad (1.136)$$

Together with a prior distribution over $\vec{\theta}_t$, $P(\vec{\theta}_t)$ and using Bayes' theorem, we can express the posterior distribution over $\vec{\theta}_t$ given a data d as

$$P(\vec{\theta}_t | d) = \mathcal{N} P(\vec{\theta}_t) \exp \left[\langle h(\theta_t) | d \rangle - \frac{1}{2} \langle h(\theta_t) | h(\theta_t) \rangle \right] \quad (1.137)$$

where $\langle d | d \rangle$ is a constant for a given data d and is reabsorbed in \mathcal{N} . In principle, you can evaluate the posterior $P(\vec{\theta}_t | d)$ at various values of $\vec{\theta}_t$. However, for a typical binary black hole (BBH) system, $\vec{\theta}_t$ belongs to a fifteen-dimensional parameter space, which presents significant computational challenge when evaluating the posterior. In Section [parameter estimation], we have explored several efficient techniques for estimating posterior distributions. This thesis primarily focuses on a rapid parameter estimation technique developed to address the computational challenges, especially for low-mass binaries like binary neutron stars (BNS), where the in-band signal has an extended duration. For now, let's focus on extracting essential information (an estimator) from the posterior, such as the maximum probable value denoted as $\hat{\vec{\theta}}$, and the corresponding estimation uncertainties or errors. An ideal estimator should adhere to fundamental criteria, including consistency, unbiasedness, efficiency, and robustness. One reasonable choice is the maximum likelihood estimator ($\hat{\vec{\theta}}_{ML}$), which is a value of $\vec{\theta}$ that maximizes the likelihood. When assuming a flat prior distribution, maximizing the posterior $P(\vec{\theta}_t | d)$ is equivalent to

maximizing the likelihood $p(d | \vec{\theta}_t)$. By taking the natural logarithm of Eq. (1.136) and omitting the constant terms, we obtain:

$$\ln P(d | \vec{\theta}_t) = \langle h(\theta_t) | d \rangle - \frac{1}{2} \langle h(\vec{\theta}_t) | h(\vec{\theta}_t) \rangle \quad (1.138)$$

Maximizing the above equation amounts to taking its derivative and setting it to zero,

$$\left\langle \frac{\partial h(\vec{\theta}_t)}{\partial \vec{\theta}_i} | d \right\rangle - \left\langle \frac{\partial h(\vec{\theta}_t)}{\partial \vec{\theta}_i} | h(\vec{\theta}_t) \right\rangle = 0 \quad (1.139)$$

and the corresponding errors, denoted as $\Delta\vec{\theta}_i$, can be estimated by using the width of the posterior distribution $P(\vec{\theta}_t | d)$. Interestingly, it is found that $\hat{\theta}_{\text{ML}}$ is identical to the $\vec{\theta}$ that maximizes ρ_m in the matched-filtering process. To illustrate this, let's represent a generic template as $h(t; \vec{\theta}) = Ah_0(t; \vec{\lambda})$, where the amplitude A depends on certain parameters, while the remaining parameters are represented by $\vec{\lambda}$. Substituting it in the Eq. (1.138) and maximizing it with respect to A , we find

$$\hat{A}_{\text{ML}}(d) = \frac{\langle h_0(\vec{\lambda}) | d \rangle}{\langle h_0(\vec{\lambda}) | h_0(\vec{\lambda}) \rangle}. \quad (1.140)$$

and the corresponding maximum likelihood is

$$\ln P(d | \vec{\lambda}) = \frac{1}{2} \frac{\langle h_0(\vec{\lambda}) | d \rangle^2}{\langle h_0(\vec{\lambda}) | h_0(\vec{\lambda}) \rangle} \quad (1.141)$$

It can be further maximized with respect to $\vec{\lambda}$ and is equivalent to maximizing the overlap between d with $h_0(\vec{\lambda})/\langle h_0(\vec{\lambda}) | h_0(\vec{\lambda}) \rangle^{1/2}$ which is the matched-filtering procedure we discussed earlier. For the signal with very high SNRs, the corresponding errors of the estimated parameters are expected to be small so that we can expand $h(\vec{\theta}_i)$ around some estimate, say $\hat{\theta}_{\text{ML}}$ and the Eq. (1.137), takes a simple form,

$$P(\vec{\theta} | d) = \mathcal{N} \exp \left[-\frac{1}{2} \Gamma^{ij} \Delta\vec{\theta}_i \Delta\vec{\theta}_j \right], \quad (1.142)$$

where

$$\Gamma^{ij} = \left\langle \frac{\partial}{\partial \vec{\theta}_i} \frac{\partial h}{\partial \vec{\theta}_j} | h - d \right\rangle + \left\langle \frac{\partial h}{\partial \vec{\theta}_i} | \frac{\partial h}{\partial \vec{\theta}_j} \right\rangle, \quad (1.143a)$$

$$\simeq \left. \left\langle \frac{\partial h}{\partial \vec{\theta}_i} | \frac{\partial h}{\partial \vec{\theta}_j} \right\rangle \right|_{\vec{\theta}_i = \hat{\theta}_{\text{ML}}} \quad (1.143b)$$

where we arrive at the Eq. (1.143b) from Eq. (1.143a) by using a large SNR limit and ignoring the first term in Eq. (1.143a). Γ^{ij} in the above equation is known as the Fisher information matrix. The inverse of the Fisher information matrix gives the estimate of the average errors over the estimated parameters, $(\Gamma^{-1})_{ij} = \langle \Delta\vec{\theta}_i \Delta\vec{\theta}_j \rangle$.

Matched filtering statistics After detecting a gravitational wave at a specific SNR, denoted as ρ_m , it is essential to assess the statistical significance of this detection, which relies on the statistical properties of the noise. When a signal is buried within Gaussian noise, the presence of any triggers due to this noise can be eliminated by setting a high threshold for ρ_m . However,

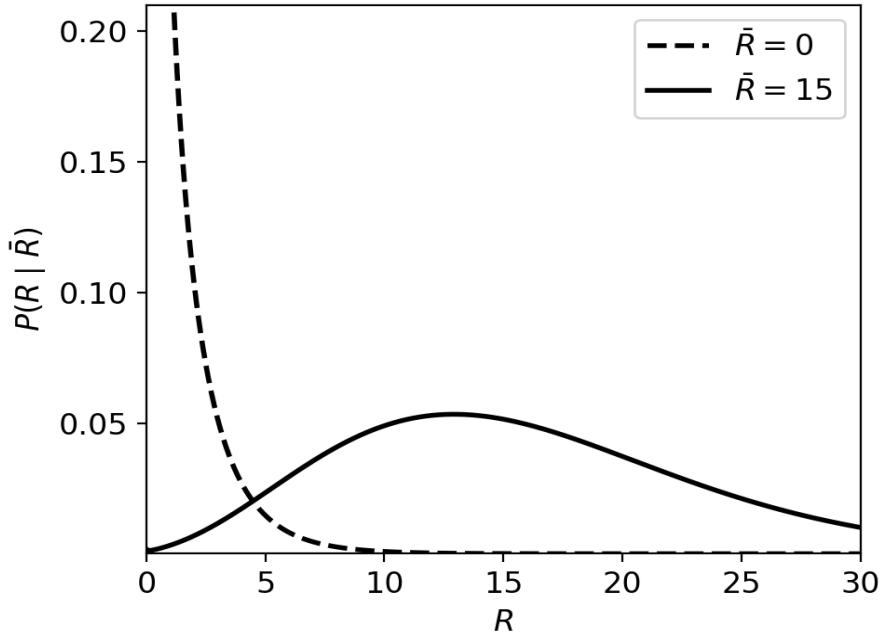


Figure 1.10: Probability distribution of R for $\bar{R} = 0$ and $\bar{R} = 15$ (reproduced from [7])

real-world scenarios often involve non-Gaussian noise, which possesses heavy tails, even at high ρ_m values, making it challenging to distinguish genuine signals from spurious non-Gaussian features. This challenge emphasizes the need for multi-detector networks to perform coincidence tests, helping mitigate issues arising from non-Gaussian noise. Recalling the SNR ρ_m , which represents the ratio of the expected value of \hat{d} when a signal is present to the root mean square (RMS) value of \hat{d} when the signal is absent, we can express \hat{d} from Equation (1.130). Introducing $\rho = \hat{d}/N$, where N is the RMS value of \hat{d} when the signal is absent, we find that $\rho_m = \langle \rho \rangle$. From Equation (1.130), it's evident that ρ is a random variable following a standard normal distribution when the signal is absent,

$$P(\rho | h = 0)d\rho = \frac{1}{\sqrt{2\pi}} \exp^{-\rho^2/2} d\rho \quad (1.144)$$

whereas in the presence of a GW signal with SNR equal to $\bar{\rho}$, $\rho = \bar{\rho} + \hat{n}/N$, where $\hat{n} = \int dt n(t)K(t)$ which implies

$$P(\rho | \bar{\rho})d\rho = \frac{1}{\sqrt{2\pi}} \exp((\rho - \bar{\rho})^2/2) d\rho \quad (1.145)$$

Defining $R^3 \equiv \rho^2$, we can write

$$P(R | \bar{R})dR = \frac{1}{\sqrt{2\pi R}} \exp[-(R + \bar{R})/2] \cosh[\sqrt{R\bar{R}}] dR. \quad (1.146)$$

As illustrated in Figure 1.10, it becomes evident that we can differentiate a genuine GW signal from fluctuations caused by Gaussian noise by setting a threshold on the parameter R at a designated value denoted as R_t . This threshold should be chosen sufficiently far into the tail of

³This is equivalent to SNR in energy so that R is always positive.

the distribution, ensuring the retention of a substantial portion of the true signals. At the same time, there also exists a false alarm probability,

$$\begin{aligned} p_{\text{FA}} &= \int_{R_t}^{\infty} dR P(R \mid \bar{R} = 0) \\ &= 2 \int_{\rho_t}^{\infty} d\rho \exp^{-\rho^2/2} \end{aligned} \quad (1.147)$$

and a false dismissal probability,

$$p_{\text{FD}} = \int_0^{R_t} dR P(R \mid \bar{R}) \quad (1.148)$$

The threshold R_t can be fixed by setting a maximum false alarm rate we are willing to risk. We can generalize Eq. (1.146), extending it for cases when we have multiple detectors and their corresponding SNRs, each following a Gaussian distribution such that $\rho^2 = \rho_1^2 + \rho_2^2 + \dots + \rho_n^2$, one can show that

$$P_n(R \mid \bar{R})dR = \frac{1}{\sqrt{2\pi}\bar{R}} \left(\frac{R}{\bar{R}} \right)^{(n-2)/4} \exp[-(R + \bar{R})/2] I_{n/2-1}[\sqrt{R\bar{R}}]dR. \quad (1.149)$$

where $I_{(\cdot)}$ modified bessel function of first kind. It is a non-central χ^2 distribution with n -degrees of freedom. In this section, we shall discuss how we can detect a coalescence of compact binaries.

Compact binary coalscences

The merger of compact binary systems like BBH, BNS, and neutron star-black hole binaries (NSBH) is particularly favorable for broad-band GW detectors [12, 31] because these signals spend a substantial number of cycles within the detector's sensitive frequency band. Of course, it crucially depends on the component masses of the binary. Matched filtering, when used effectively, can prove invaluable for extracting these signals, even when they are deeply embedded in a noisy background. To achieve this, it's crucial to have a precise understanding of the waveform associated with specific binary parameters. Using Eq. (1.114) and Eq. (1.77), we can write $h(t)$ as

$$h(t) = A \left(\frac{\pi f_{\text{gw}}(t)}{c} \right)^{2/3} \cos[\Phi(f_{\text{gw}}(t)) + \phi] \quad (1.150)$$

where ϕ is some constant phase which is related to Φ at colascence and $A = \sqrt{A_+^2 + A_\times^2}$. There are some extrinsic parameters, such as time shifts and phase, which can be eliminated. To see this, observe that the frequency domain GW signal at some time of coalescence t_c can be calculated by simply using a time-translation, $\tilde{h}(f; \vec{\theta}) \exp(i2\pi f t_c)$, where $\tilde{h}(f; \vec{\theta})$ is the Fourier transform of the GW signal $h(t; \vec{\theta}, t_c = 0)$. Using the definition of inner product, we find

$$\langle h(\vec{\theta}, t^*) \mid d \rangle = 4\text{Re} \int_0^{\infty} df \frac{\tilde{h}^*(f; \vec{\theta}) \tilde{d}(f)}{S_n(f)} \exp(i2\pi f t_0) \quad (1.151)$$

So, in a single Fast Fourier Transform (FFT), we can evaluate $\langle h(\vec{\theta}, t^*) \mid d \rangle$ at all possible time shifts, implying we can immediately locate the t_c corresponding to highest matched-filter SNR as evident from the Fig. 1.11. Apart from the time of coalescence, we can also analytically maximize

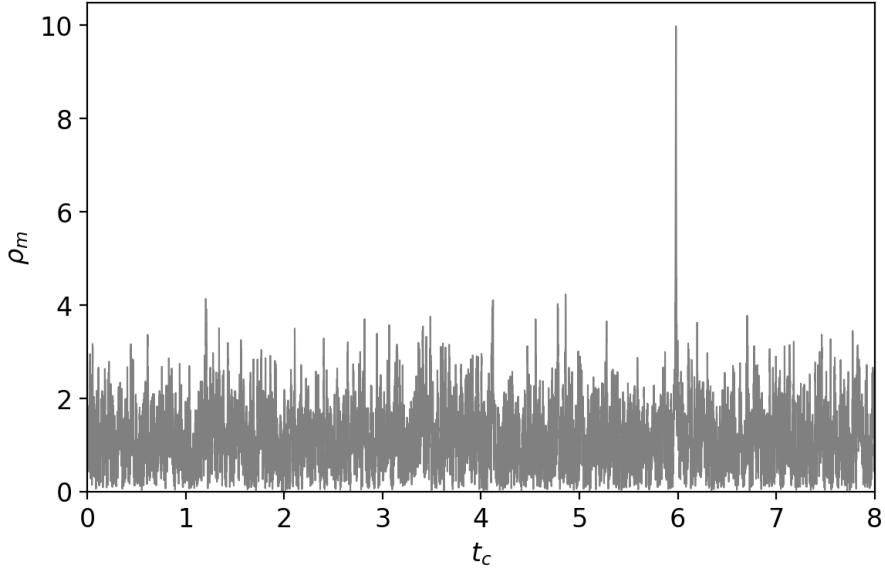


Figure 1.11: The matched-filter SNR for a BBH signal with component masses $10 M_\odot$ each, at a distance of 700 Mpc injected into the noise of the LIGO detector. The time of coalescence ($t_c = 6$) is characterized by the peak of the SNR (made using PyCBC [15, 16]).

$\ln \mathcal{L}$ over A (see Sec. 1.1.7) and ϕ . Additionally, one can show that the SNR average over the sky angles and inclination is

$$\langle \rho \rangle = \frac{2}{5} \left(\frac{5}{6} \right)^{1/2} \frac{1}{\pi^{2/3}} \frac{c}{r} \left(\frac{G\mathcal{M}}{c^3} \right)^{5/6} \times \sqrt{\left[\int_0^{f_{\max}} df \frac{f^{-7/3}}{S_n(f)} \right]}. \quad (1.152)$$

This formula can be inverted to calculate the maximum observable distance, r , at which a binary coalescence can be detected with a specific angle-inclination-averaged SNR, denoted as $\langle \rho \rangle$. This relationship between distance and SNR implies that a greater number of cycles, \mathcal{N}_c , that a GW signal spends within the detector's sensitivity band results in a higher accumulated SNR. Additionally, it can be demonstrated that the relative error in measuring the chirp mass \mathcal{M} is inversely proportional to the number of in-band cycles, indicating that events with a higher number of cycles will have lower chirp mass errors compared to those with fewer in-band cycles. In the following section, we will discuss the GW interferometers and provide a brief overview of current and future GW detectors.

1.1.8 GW Detectors: Interferometers

The concept of employing interferometers for gravitational wave (GW) detection was initially introduced by M. Gertsenshtein and V.I. Pustovoit in 1962, and it was subsequently advanced by R. Weiss, R. Drever, and other researchers. In this section, we will discuss the underlying principles of interferometer-based GW detectors and provide a concise survey of various GW detectors.

Michelson interferometer

The Michelson interferometer [37], employed by Michelson and Morley in their renowned 1887 experiment [38] to investigate the presence of ether, is a specific type of interferometer. This instrument, as depicted in Fig. 1.12, was devised to detect changes in the time taken by light to travel through its arms. It is comprised of a monochromatic light source (typically a laser) that emits light directed through a beam-splitter, dividing it into two beams of equal amplitude. These two beams travel along arms that are oriented perpendicular to each other. At the end of each arm, a mirror with total reflectivity redirects the light beams back towards the beam splitter, where they are recombined. A portion of this recombined beam travels to a photodetector, which quantifies its intensity, while the remaining part goes back to the light source. Let's denote the

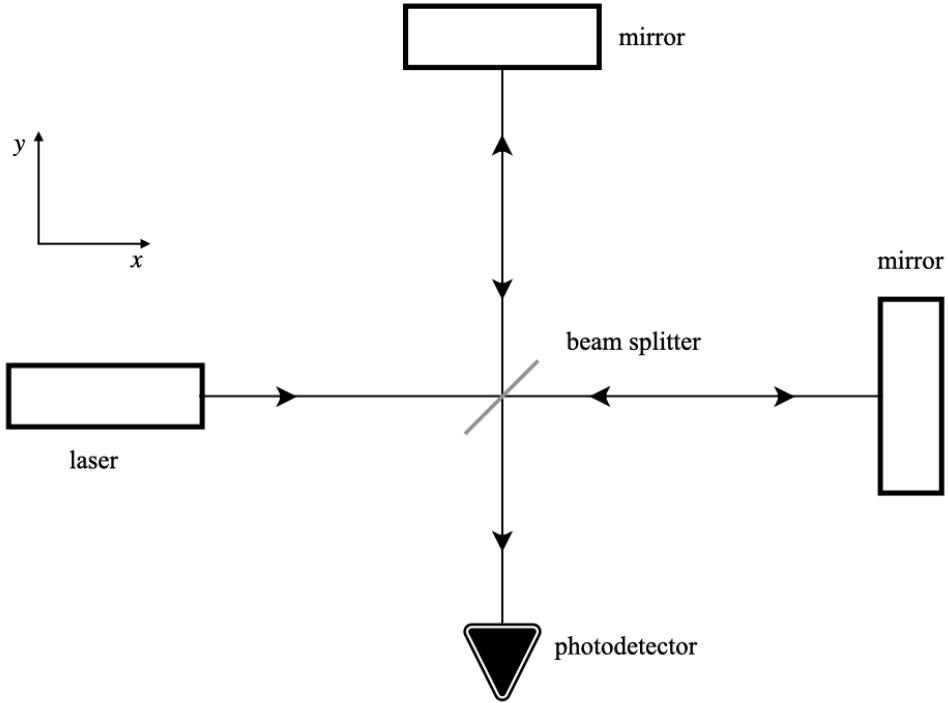


Figure 1.12: A simple setup of a Michelson interferometer (reproduced from [7]).

length of the two arms by L_x and L_y , respectively. There exist an Electric field of the input laser light whose spatial component is given by

$$E_0 e^{-i\omega_L t + i\vec{k}_L \cdot \vec{x}} \quad (1.153)$$

where ω_L is the laser's frequency, $k = \omega_L/c$ represents wave number and $\lambda_L = 2\pi/k_L$ defines the wavelength of the laser light. Consider a photon arriving at the beam splitter at some initial time t_0 . The part of the Electric field that travels through x -arm and bounces off the mirror at the end of the arm and returns back to the beam splitter at time $t = t_0 + 2L_x/c$, while the beam which traveled along the y -arm arrived back at the beam splitter at $t' = t_0 + 2L_y/c$. Defining $\vec{x} = 0$ at the beam splitter, the two electric fields which recombine at the beam splitter are given

by

$$E_1 = -\frac{1}{2}E_0 e^{-i\omega_L t + 2ik_L L_x} \quad (1.154a)$$

$$E_2 = \frac{1}{2}E_0 e^{-i\omega_L t + 2ik_L L_y} \quad (1.154b)$$

So that the recombined electric field is

$$E_{\text{out}} = E_1 + E_2 = -iE_0 e^{-i\omega_L t + ik_L(L_x + L_y)} \sin[k_L(L_x - L_y)] \quad (1.155)$$

and the corresponding power measured by the photodetector is given by

$$|E_{\text{out}}|^2 = E_0^2 \sin^2[k_L(L_y - L_x)]. \quad (1.156)$$

The above equation indicates that any change in the arm's length causes a corresponding variation of the power detected at the photodetector. Now, we turn to discuss the interaction of the detector with GWs within the TT gauge. As discussed in earlier sections, positions are determined by freely falling objects in the TT gauge. Consequently, the coordinates of these freely falling objects remain constant even in the presence of GWs ⁴. This implies that the GWs do not affect the mirror and beam splitter coordinates. For the sake of simplicity, we will assume only a "+" polarization of the GW originating from the z -direction. In the plane perpendicular to the direction of propagation, we can express the plus polarization as $h_+(t) = h_0 \cos(\omega_{\text{gw}} t)$. For a photon following a null geodesic path, we have the following expression:

$$dx = \pm cdt \left[1 - \frac{1}{2}h_+(t) \right], \quad (1.157)$$

where the plus sign denotes the time it takes for light to travel from the beam splitter to the mirror, while the negative sign represents the time for the return journey. If we consider a photon departing from the beam splitter at time t_0 and arriving at the mirror at time t_1 , and then returning to the beam splitter at time t_2 , we can determine t_2 in terms of t_0 and $h_+(t)$ using the equation above as:

$$\begin{aligned} \Delta t \equiv t_2 - t_0 &= \frac{2L_x}{c} + \frac{1}{2} \int_{t_0}^{t_2} dt' h_+(t') \\ &= \frac{2L_x}{c} + \frac{L_x}{c} h(t_0 + L_x/c) \frac{\sin(\omega_{\text{gw}} L_x/c)}{\omega_{\text{gw}} L_x/c}. \end{aligned} \quad (1.158)$$

Similarly, for the y -arm, the above expression becomes,

$$\Delta t \equiv t_2 - t_0 = \frac{2L_y}{c} + \frac{L_y}{c} h(t_0 + L_y/c) \frac{\sin(\omega_{\text{gw}} L_y/c)}{\omega_{\text{gw}} L_y/c}. \quad (1.159)$$

⁴It is important to note that while the mirrors are not in true free fall and are instead held in place by suspensions to counteract the Earth's gravity, these forces are essentially static in comparison to the time-scales of the GWs under study. Nonetheless, the mirrors can be considered to be in free fall in the horizontal plane.

Inverting the above two equations in the first order in h_0 to get the corresponding value of t_0 for a given value of $t \equiv t_2$, we find

$$t_0^{(x)} = t - \frac{2L_x}{c} - \frac{L_x}{c} h(t - L_x/c) \sin c(\omega_{\text{gw}} L_x/c), \quad (1.160\text{a})$$

$$t_0^{(y)} = t - \frac{2L_y}{c} - \frac{L_y}{c} h(t - L_y/c) \sin c(\omega_{\text{gw}} L_y/c). \quad (1.160\text{b})$$

The corresponding electric fields are

$$E^{(x)}(t) = -\frac{1}{2} E_0 e^{-i\omega_L(t-2L/c)+i\phi_0+i\Delta\phi_x(t)}, \quad (1.161\text{a})$$

$$E^{(y)}(t) = -\frac{1}{2} E_0 e^{-i\omega_L(t-2L/c)-i\phi_0+i\Delta\phi_y(t)} \quad (1.161\text{b})$$

where $L = (L_x + L_y)/2$, $\phi_0 = k_L(L_x - L_y)$, and $\Delta\phi_x$ given by

$$\begin{aligned} \Delta\phi_x(t) &= h_0 k_L L \sin c(\omega_{\text{gw}} L/c) \cos(\omega_{\text{gw}}(t - L/c)), \\ &\equiv |\Delta\phi_x| \cos(\omega_{\text{gw}} t + \alpha), \end{aligned} \quad (1.162)$$

with $\alpha = -\omega_{\text{gw}} L/c$.

The total phase difference caused by GWs in the interferometer can be shown as $\Delta\phi_{\text{Mich}} \equiv \Delta\phi_x - \Delta\phi_y = 2\Delta\phi_x$ where $\Delta\phi_y = -\Delta\phi_x$ and corresponding total electric field is given by

$$E_{\text{tot}}(t) = -i E_0 e^{-i\omega_L(t-2L/c)} \sin[\phi_0 + \Delta\phi_x(t)]. \quad (1.163)$$

where ϕ_0 is a parameter chosen by the experimenter. In the limit $\omega_{\text{gw}} L/c \ll 1$ which is equivalent to $\lambda_{\text{gw}} \gg L$, Eq. (1.162) becomes $\Delta\phi_x(t) \simeq h(t - L/c)k_L L$ which implies that in this limit, GWs effect on the phase shift is equivalent to a variation in $L_x - L_y$ and can be written as

$$\frac{\Delta(L_x - L_y)}{L} \simeq h(t - L/c) \quad (1.164)$$

From the Eq. (1.163), we find the total power observed at the photodetector is

$$\begin{aligned} P &= P_0 \sin^2[\phi_0 + \Delta\phi_x(t)] \\ &= \frac{P_0}{2} (1 - \cos[2\phi_0 + \Delta\phi_{\text{Mich}}(t)]). \end{aligned} \quad (1.165)$$

From the above equation, it is evident that the $\Delta\phi_{\text{Mich}}$ is clearly dependent on the L apart from h_0 . So the optimal length of the arms is $\omega_{\text{gw}} L/c = \pi/2$ or in terms of f_{gw} given by

$$L \simeq 750 \text{ km} \left(\frac{100 \text{ Hz}}{f_{\text{gw}}} \right). \quad (1.166)$$

In the case of this arm length L , the time shift caused by GW consistently has the same sign throughout the entire round trip within the arm. This leads to an accumulation of the effect. However, in the case of longer arms, the amplitude of GW causes the sign to reverse during the round trip, which results in the cancellation of the phase shift that the light has accumulated. To incorporate this optimal path length into an interferometer of practical size, a Fabry-Perot cavity [39, 40] is used, which we discuss in the next section.

Fabry-Perot cavity

From the Equation (1.166), it becomes evident that in order to detect GWs with a frequency on the order of 100 Hz, it would necessitate arm lengths (denoted as L) extending several hundred kilometers. However, achieving such extensive arm lengths for Earth-based interferometers is both practically and financially infeasible. The solution lies in attaining an optimal path length of 750 km using an interferometer with arms that are only a few kilometers long. For example, the LIGO interferometer features arms with a length of 4 km, while Virgo's arms are 3 km long. This is made possible through the utilization of a Fabry-Perot (FP) cavity. Let's briefly explore how the FP cavity enables us to reach the required optimal path length.

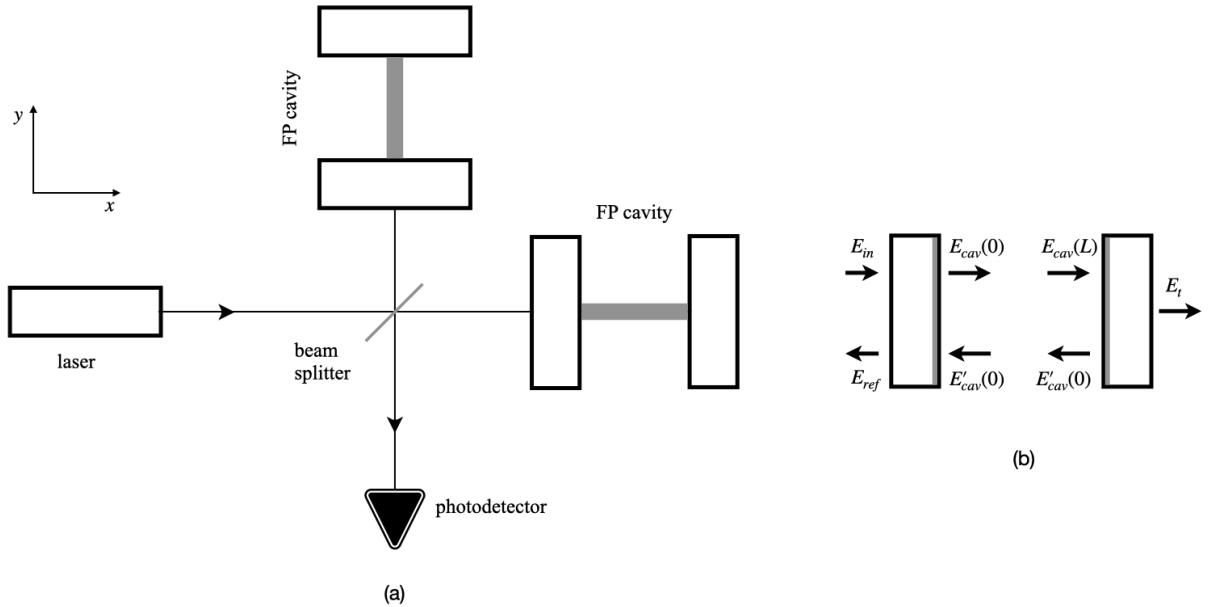


Figure 1.13: Left panel: A simple setup of a Michelson interferometer with Febry-Perot cavities. Right panel: A schematic of Febry-Perot cavity (reproduced from [7])

An FP cavity is constructed with two parallel mirrors, both featuring high-reflectivity coatings on their interior surfaces, as depicted in the second panel of Figure 1.13. The incoming electric field E_{in} enters from the left side of the cavity, with a portion being reflected by the coating while the remainder is transmitted. The transmitted portion $E_{cav}(0)$ traverses through the cavity, where it experiences partial reflection and partial transmission. The reflected portion $E_{cav}(L)$ journeys back to the initial mirror, once again undergoing partial reflection and transmission and so on. The total fields for reflection and transmission inside the cavity are composed of multiple beams resulting from numerous reflections. Using reflection and transmission coefficients denoted as r_1 and t_1 for the first mirror and r_2 and t_2 for the second mirror, we can express $E_{cav}(0)$ in terms of E_{in} and $E'_{cav}(0)$ as follows:

$$E_{cav}(0) = t_1 E_{in} + r'_1 E'_{cav}(0), \quad (1.167a)$$

$$E_{ref}(0) = r_1 E_{in} + t'_1 E'_{cav}(0). \quad (1.167b)$$

where $E_{in} = E_0 e^{-i\omega_L t}$. We can write similar equations at the second mirror as

$$E_t = t_2 E_{cav}(L), \quad (1.168a)$$

$$E'_{cav}(L) = -r'_2 E_{cav}(L). \quad (1.168b)$$

In the above equations, we can set $r'_{1(2)} = -r_{1(2)}$ and $t'_{1(2)} = t_{1(2)}$ ⁵. The solution within the cavity consists of plane waves, which means that the cavity fields at positions $x = L$ and $x = 0$ at the same time are related by the following equation:

$$E_{cav}(L) = e^{ik_L L} E_{cav}(0), \quad (1.169a)$$

$$E'_{cav}(L) = e^{-ik_L L} E'_{cav}(0) \quad (1.169b)$$

Solving the above equations for E_{ref} , E_t , and $E_{cav}(0)$, we find

$$E_{ref} = E_0 e^{-i\omega_L t_0} \frac{r_1 - r_2(1 - p_2)e^{2ik_L L}}{1 - r_1 r_2 e^{2ik_L L}}, \quad (1.170a)$$

$$E_t = E_0 e^{-i\omega_L t_0} \frac{t_1 t_2 e^{ik_L L}}{1 - r_1 r_2 e^{2ik_L L}}, \quad (1.170b)$$

$$E_{cav}(0) = E_0 e^{-i\omega_L t_0} \frac{t_1}{1 - r_1 r_2 e^{2ik_L L}} \quad (1.170c)$$

where p_1 represents losses in the first mirror and is related to r_1 and t_1 by $r_1^2 + t_1^2 = 1 - p_1$. From the above equations, we can find the power of the transmitted field P_t as

$$P_t = |E_t|^2 = E_0 \frac{t_1^2 t_2^2}{1 + (r_1 r_2)^2 - 2r_1 r_2 \cos(2k_L L)} \quad (1.171)$$

The separation between the peaks in the function above is referred to as the free spectral range of the cavity and is quantified as $\Delta\omega_L = \pi c/L$. Meanwhile, the full width of the peaks at half maximum (FWHM) is expressed as $\delta\omega_L = c(1 - r_1 r_2)/L\sqrt{r_1 r_2}$. We can define the finesse of the cavity, denoted as \mathcal{F} , as the ratio of the free spectral range to the FWHM,

$$\mathcal{F} = \frac{\pi\sqrt{r_1 r_2}}{1 - r_1 r_2} \quad (1.172)$$

We can calculate another important quantity called storage time, denoted as τ_s , in the context of the FP. It is defined as the average duration a photon remains inside the cavity. It can be expressed in terms of the finesse as follows:

$$\tau_s \simeq \frac{L \mathcal{F}}{c \pi} \quad (1.173)$$

In fact, the intensity of light that comes out from the first mirror after undergoing multiple round trips is dependent on τ_s and decreases with time. From Eq. (1.170a), we can write down E_{ref} at resonances as

$$E_{ref} = E_0 e^{-i\omega_L t_0} \frac{r_1 - r_2(1 - p_1)}{1 - r_1 r_2} \quad (1.174)$$

⁵This can be proved by assuming no gap between the two mirrors. For a perfectly reflecting mirror, $r'_1 = -1$ when the reflection occurs from a less dense to a more dense medium while $r = 1$ when it happens from the denser side to the less dense side.

where resonances are defined to be the conditions when $2k_L L = 2\pi n$ where $n = 0, \pm 1, \pm 2, \dots$. This condition implies that the beams bouncing back and forth inside the cavity undergo constructive interference, resulting in a significantly amplified field within the cavity. Note that there is no reflected light if $r_1 = r_2(1 - p_1)$. The light immediately reflected back undergo destructive interference with the light that gets reflected after going through either one or multiple round trips within the cavity, causing all the incident light to leak out from the second mirror. This scenario is referred to as critical cavity coupling, and it's highly undesirable in the context of a GW interferometer because we are interested in the reflected light. Now, from Eq. (1.172), we can write $r_1 r_2 = 1 - \pi/\mathcal{F} + \mathcal{O}(\pi^2/\mathcal{F}^2)$ and introduce coupling rate σ defined as $\sigma = p\mathcal{F}/\pi$ where $p = 1 - (1 - p_1)r_2^2$ which encapsulates the losses from both mirrors. Using these expressions, we can write

$$\frac{r_1 - r_2(1 - p_1)}{1 - r_1 r_2} = \frac{\sigma - 1}{r_2} \quad (1.175)$$

From the above equation, we can see that $\sigma = 1$ corresponds to critical coupling. For $0 < \sigma < 1$,

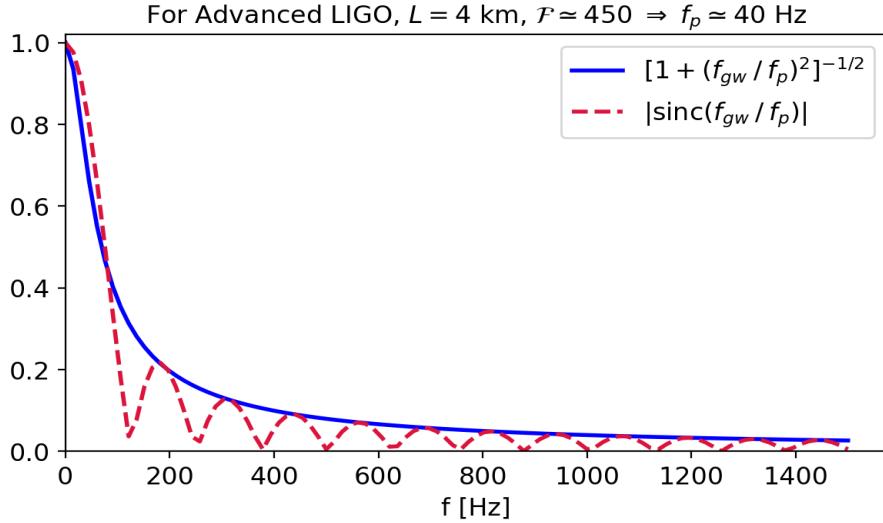


Figure 1.14: Plot of the factors affecting the phase of the FP cavity and Michelson interferometer. It is clear that the sinc function oscillates rapidly beyond a certain frequency, canceling the phase contributions in contrast to the other factor.

the cavity is overcoupled, equivalent to having a higher sensitivity to changes of $2k_L L$, which is crucial for the phase of the reflected field. For the arms of Advanced LIGO [41] and Advanced Virgo [31], $p \sim 2 \times 10^{-5}$ with \mathcal{F} to be ≈ 450 , giving $\sigma \sim 10^{-3}$, which indicates that LIGO and Virgo's cavities are well overcoupled. Without going into the technical details of the effect of GWs on FP cavities, we write down the phase shift in an FP interferometer due to GWs as

$$|\Delta\phi_{\text{FP}}| \simeq h_0 \frac{4\mathcal{F}}{\pi} k_L L \frac{1}{\sqrt{1 + (f_{\text{gw}}/f_p)^2}}. \quad (1.176)$$

where f_p is called pole frequency and defined as $f_p \equiv 1/4\pi\tau_s \simeq c/4\mathcal{F}L$ and the above expression

is valid if $\omega_{\text{gw}}L/c \ll 1$ or equivalently,

$$\begin{aligned} f_{\text{gw}} &\ll \frac{c}{2\pi L}, \\ &\simeq 12\text{kHz} \left(\frac{4\text{km}}{L} \right). \end{aligned} \quad (1.177)$$

In the Fig. 1.14, $[1 + (f_{\text{gw}} / f_p)^2]^{-1/2}$ (see Eq. (1.176)) is compared to $|\text{sinc}(f/f_p)|$. The equivalent length L_{Mich} of the Michelson interferometer needed to match sensitivity of a FP cavity with length L and Finesse \mathcal{F} is given by $L_{\text{Mich}} = 2\mathcal{F}L/\pi$. To give a comparison, for an ordinary Michelson interferometer to achieve similar sensitivity as Advanced LIGO [41] (an FP cavity-based interferometer) with $L = 4$ km, $\mathcal{F} \simeq 450$, we would need a whopping $L_{\text{Mich}} \simeq 1100$ km long arms-lengths.

In summary, the use of an FP cavity significantly enhances detector's sensitivity (by a factor of $(2/\pi)\mathcal{F}$) to a phase shift. This enhancement allows us to achieve a response to GWs equivalent to that of a Michelson interferometer with arms spanning on the order of hundreds of kilometers, which is optimal for gravitational waves with frequencies $f_{\text{gw}} = \mathcal{O}(10^2)$ Hz. We can achieve this by replacing the long arms with FP cavities of just a few kilometers in length and finesse of $\mathcal{F} = \mathcal{O}(10^2)$. In the next section, we will provide a brief overview of the various noise sources in the LIGO-Virgo interferometers.

Noise sources

As discussed in the previous section, we measure the phase shift, which encodes the signatures of the GWs. We expect to have a GW amplitude of $h_0 \sim \mathcal{O}(10^{-21})$, the corresponding displacement $\Delta L = h_0 L/2$ of the mirrors for interferometer with $L = 4$ km comes out to be $\mathcal{O}(10^{-18})$ m. It amounts to a Michelson phase shift of $\Delta\phi_{\text{Mich}} \sim \mathcal{O}(10^{-11})$ rad but for a FP cavity with $\mathcal{F} \simeq 450$, we get the the FP phase shift to be $\phi_{\text{FP}} \sim \mathcal{O}(10^{-8})$ rad. Unfortunately, the measurement of such minute phase shifts is influenced by various noise sources, which impose limitations on the interferometer's sensitivity. This strain sensitivity is quantified by $S_n^{1/2}(f)$ (measured in units of $\text{Hz}^{-1/2}$). Some of the prominent noise sources are as follows:

- **Shot noise** This noise source arises because laser light is composed of discrete quanta known as photons. Measurement of the power of a light source is equivalent to counting the number of photons that arrive within a specific time frame. Considering an FP cavity with a finesse \mathcal{F} and length L , for an optimal orientated periodic GW source radiating at frequency f with only plus polarization, the strain sensitivity due to shot noise is given by

$$S_n(f)^{1/2} \Big|_{\text{shot}} = \frac{1}{8\mathcal{F}L} \left(\frac{4\pi\hbar\lambda_L c}{\eta P_{\text{bs}}} \right)^{1/2} \sqrt{1 + (f/f_p)^2}, \quad (1.178)$$

where P_{bs} represents the beam splitter's power after recycling, which is equivalent to CP_0 with C being the power recycling factor for the laser, P_0 is the input power from the laser, and η is the factor limiting the power of extraction of electrons at the photodiode.

- **Radiation pressure** As indicated by Equation (1.178), shot noise reduction can be achieved by either increasing P_0 or increasing the recycling factor C . However, the photons

striking the mirror and reflecting back also exert radiation pressure on the mirror, causing fluctuations in response to the number of photons reaching the mirror, ultimately resulting in mirror vibrations. This pressure grows as $\sqrt{P_{\text{bs}}}$ while shot noise decreased as $1/\sqrt{P_{\text{bs}}}$. This suggests that beyond a certain increase in P_{bs} , radiation pressure starts dominating over shot noise. The radiation pressure's strain sensitivity is given by:

$$S_n(f)^{1/2} \Big|_{\text{rad}} = \frac{16\sqrt{2}\mathcal{F}}{ML(2\pi f)^2} \sqrt{\frac{\hbar P_{\text{bs}}}{2\pi\lambda_L c}} \frac{1}{\sqrt{1 + (f/f_p)^2}}. \quad (1.179)$$

where M is the mass of the mirror.

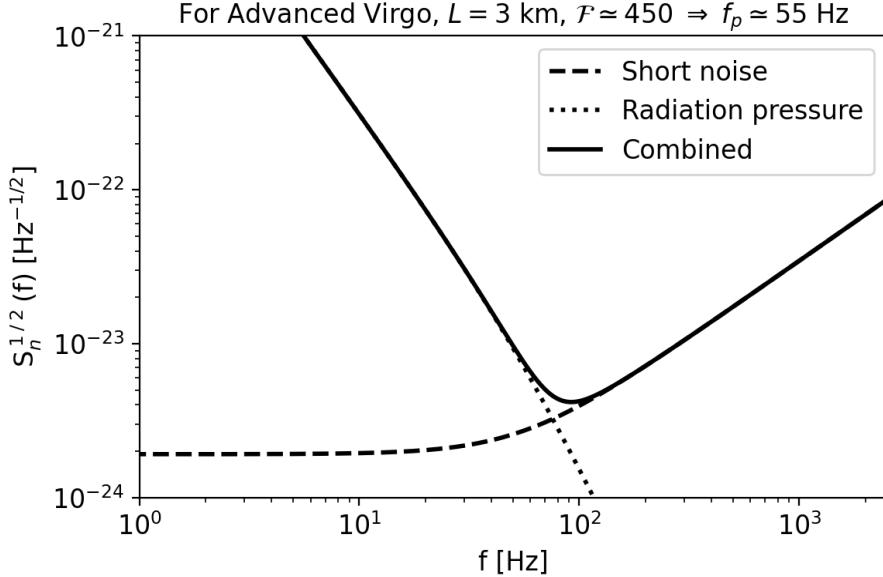


Figure 1.15: The strain sensitivity $S_n^{1/2}(f)$ representing shot noise, radiation pressure, and the combined noise.

- **The standard quantum limit** We can write the combined noise encompassing both shot noise and radiation pressure as the following: (see Fig. 1.15):

$$S_n(f)|_{\text{opt}} = S_n(f)|_{\text{shot}} + S_n(f)|_{\text{rad}}. \quad (1.180)$$

As we discussed earlier, shot noise and radiation pressure are competing against each other. The position of an object is measured using photons, which in turn is disturbed by radiation pressure exerted by themselves. Let's rewrite the above equation as the following:

$$S_n(f)|_{\text{opt}} = \frac{1}{L\pi f_0} \sqrt{\frac{\hbar}{M}} \left[\left(1 + \frac{f^2}{f_p^2} \right) + \frac{f_0^4}{f^4} \frac{1}{1 + (f/f_p)^2} \right]^{1/2} \quad (1.181)$$

where $f_0 = \frac{8\mathcal{F}}{2\pi} \sqrt{P_{\text{obs}}/\pi\lambda_L c M}$. The above $S_n(f)|_{\text{opt}}$ can be minimized with respect to f_0 and optimal value of f_0 is one for which $S_n(f)|_{\text{shot}} = S_n(f)|_{\text{rad}}$ which is given by

$$1 + \frac{f^2}{f_p^2} = \frac{f_0^2}{f_p^2}. \quad (1.182)$$

Pluggin' it in the above equation to get the optimal value of $S_n(f)|_{\text{opt}}$, also known as the standard quantum limit (SQL),

$$S_n(f)^{1/2}|_{\text{SQL}}(f) = \frac{1}{2\pi f L} \sqrt{\frac{8\hbar}{M}}. \quad (1.183)$$

It represents the lower bound of the family of functions $S_n^{1/2}(f; f_0)|_{\text{opt}}$ parameterized by f_0

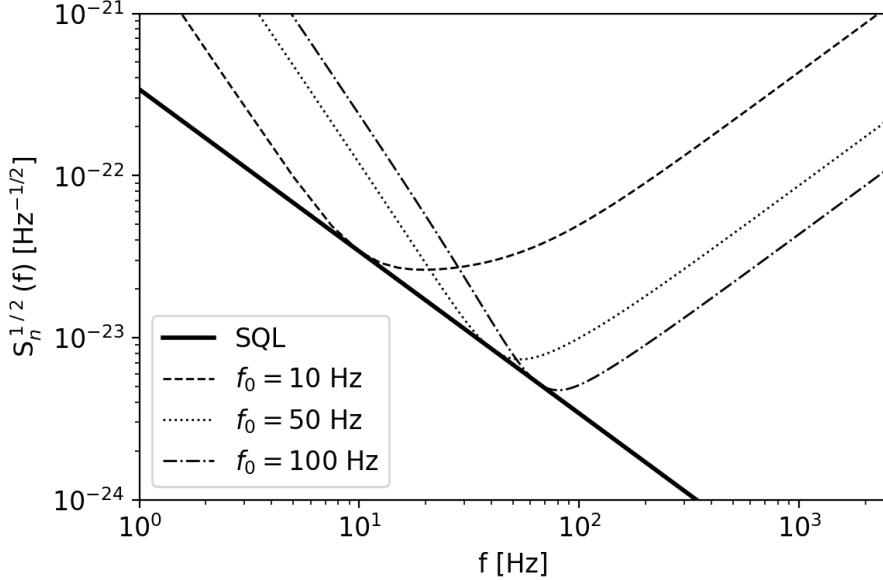


Figure 1.16: $S_n(f)_{\text{opt}}$ for different values of f_0 along with SQL.

- **Seismic, Newtonian** In addition to the aforementioned noise sources, it's important to consider those mirror movements that can be caused by factors unrelated to gravitational waves. These sources of noise fall into the category of displacement noise, and they give rise to a corresponding strain sensitivity denoted as $x(f)$, which can be described by the following equation:

$$x(f) \simeq A \left(\frac{1 \text{Hz}}{f^\nu} \right) \text{mHz}^{1/2}, \quad (1.184)$$

where index $\nu \simeq 2$ and $A \sim \mathcal{O}(10^{-7})$. Dividing the above equation by L gives the noise strain sensitivity, which is at least $\sim \mathcal{O}(10^{10})$ larger than the aimed values. This clearly indicates the need to significantly attenuate seismic noise, often accomplished through a series of pendulum systems. However, for ground-based detectors, seismic noise can only be effectively reduced for GW frequencies greater than approximately 10 Hz. Consequently, these detectors are unable to detect GWs below around 10 Hz. Newtonian noise, on the other hand, originates from time-varying Newtonian gravitational forces caused by moving objects and is induced by micro-seismic noise that couples to the interferometer's mirrors. Unfortunately, we cannot entirely eliminate Newtonian noise, as gravitational forces cannot be shielded. While it is not the primary source of noise in current GW interferometers, it may become a limiting factor in future detectors as we strive to detect even lower frequencies. In addition to these noise sources, there are other factors that can impact the

strain sensitivity of GW interferometers. For more in-depth information, interested readers can refer to the following references [7, 42].

GW detectors: Initial to advanced and beyond

Ground based detectors As previously explained, the fundamental structure of ground-based gravitational wave detectors consists of a Fabry-Perot (FP) cavity-based interferometer. There are two LIGO detectors: one situated in Livingston, Louisiana, and the other in Hanford, Washington. Each of these detectors has 4 km long arms and is separated by approximately 10 ms in terms of light travel time. The researchers and scientists who are collaborating on this



Figure 1.17: LIGO detectors [12]



Figure 1.18: Virgo detector [31]

project are members of the LIGO scientific collaboration (LSC). The aerial views of the two LIGO detectors [12] are shown in Fig. 1.17. The Virgo interferometer [31], located near Pisa, Italy, is hosted by the European Gravitational Observatory (EGO), which is a collaboration between France and Italy. It has an arm length of 3 km, which is shorter than the arm lengths of the two LIGO interferometers. An aerial view of the same is shown in Fig. 1.18. Situated near the city of Hida in Gifu Prefecture, Japan, the Kamioka Gravitational Wave Detector (KAGRA) [43, 44, 45] is another interferometer-based GW detector. It has a 3 km long arm. Figure 1.19 illustrates the target strain sensitivities of the LIGO, Virgo, and KAGRA detectors for various observing

runs. As of now, O4 is ongoing, while O5 represents forthcoming observing runs in which all the detectors aim to reach their respective theoretical design sensitivities. Another gravitational

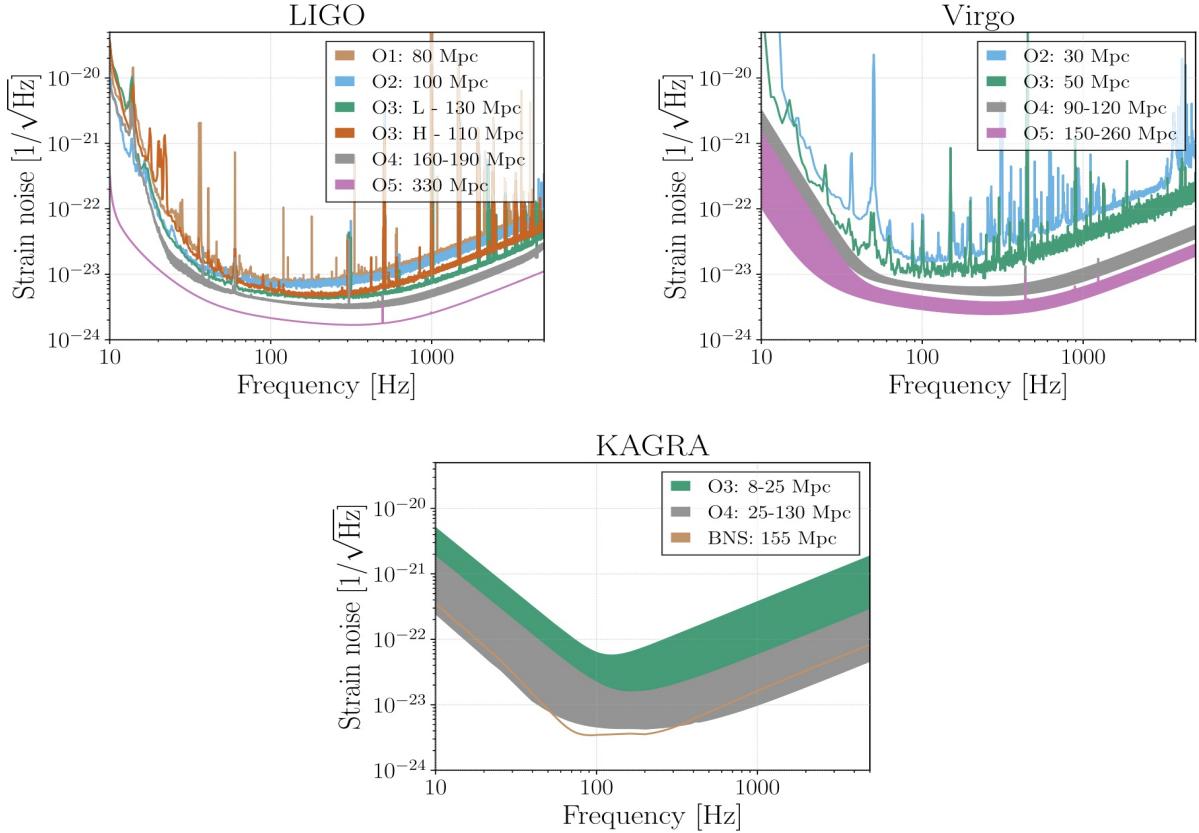


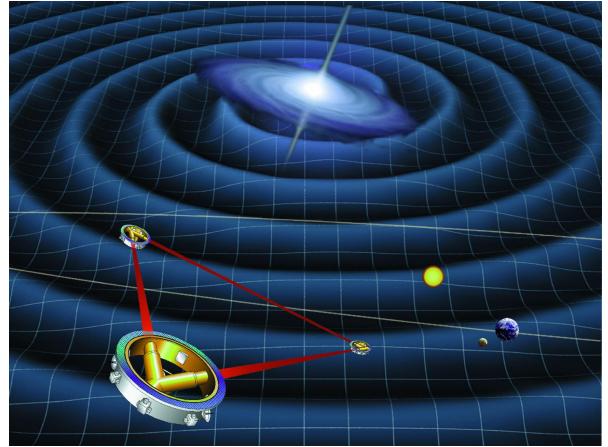
Figure 1.19: aLIGO, AdV, and KAGRA target sensitivities are shown as a function of frequency. The quoted range is for $1.4M_{\odot} + 1.4M_{\odot}$ BNS merger [46].

wave (GW) detector following the LIGO design [47, 48, 49, 46, 50] is scheduled for construction in Aundha, Maharashtra, India, as illustrated in Figure 1.20a. This facility will feature 4 km long arms and is anticipated to become part of the worldwide network of GW detectors by the year 2030. In addition to the ground-based GW detectors mentioned earlier, commonly referred to as the second generation (“2G”) detectors, there are also plans for third-generation (“3G”) detectors like the Einstein Telescope [51, 52] and Cosmic Explorer [53, 54].

Space borne detectors As previously discussed, ground-based gravitational wave detectors cannot access frequencies below 10 Hz due to the presence of seismic and Newtonian noise. To explore the lower frequency range, especially in the millihertz (mHz) region where supermassive black holes could be potential GW sources, space-based observatories are essential due to the absence of seismic noise. Several planned space-based detectors, including the Laser Interferometer Space Antenna (LISA) [55] and DECi-hertz Interferometer Gravitational-wave Observatory (DECIGO) [56, 57, 58], aim to accomplish this. LISA is a collaborative project between the European Space Agency (ESA) and the National Aeronautics and Space Administration (NASA). It consists of three spacecraft positioned in an equilateral triangle configuration with sides measuring 5 million km, and the center of this triangle is located approximately 50 million



(a) An artistic impression of LIGO-Aundha (source ~ link)



(b) An artistic impression of LISA (source ~ ESA)

km behind Earth. LISA has been optimized for detecting GW frequencies in the range of 0.1 mHz to 0.1 Hz. An artist's representation of LISA is provided in Figure 1.20b. DECIGO is also a proposed space-based GW detector, a Japanese initiative designed to be sensitive to GW frequencies between 0.1 Hz and 10 Hz. The characteristic strain sensitivity of all these detectors can be seen in Fig. 1.21, with DECIGO effectively bridging the gap between the sensitive frequency bands of LIGO and LISA.

In the next section, we shall discuss various strategies to sample probability distributions, which are the backbone of the Bayesian source reconstruction of GW sources.

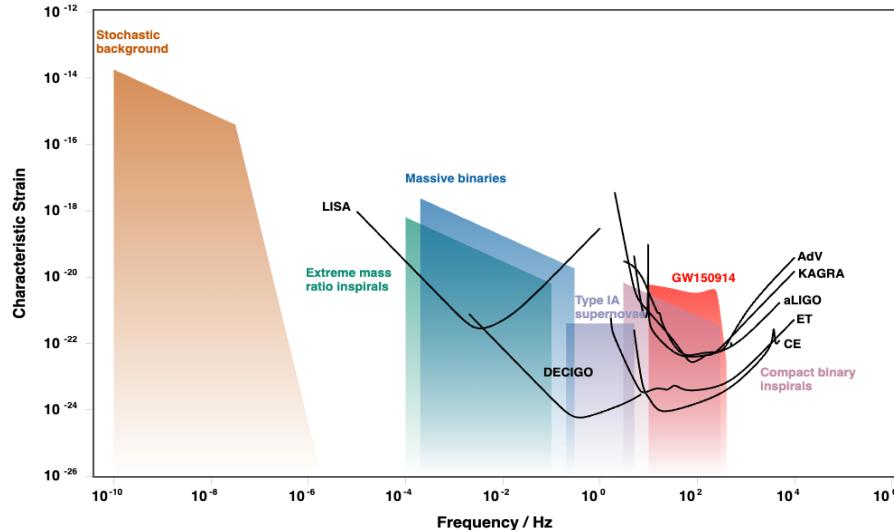


Figure 1.21: Noise strain sensitivity in various GW detectors, encompassing both ground-based and space-based observatories, is depicted. It is evident that LISA, DECIGO, and other ground-based detectors collectively span a wide spectrum of GW frequencies. This comprehensive coverage is particularly valuable for conducting multiband GW analyses [46] (made using gwplotter as explained in [59]).

1.2 Parameter estimation

1.2.1 Inverse problems

One of the objectives in the scientific investigation of a particular problem is to establish a connection between the physical parameters defining a model, denoted as \mathcal{X} , and the observations in the form of data, denoted as d , obtained through experimental procedures. Assuming an adequate understanding of the underlying physics, it is possible to define a function \mathcal{G} in such a way that

$$\mathcal{G}\mathcal{X} = d \quad (1.185)$$

Typically, real observations are invariably affected by noise, stemming from various factors like unaccounted influences on instrument readings, numerical approximations, and more. It is reasonable to consider the data as a combination of pure, noise-free observations denoted as d_{true} , accompanied by a noise component ς

$$\begin{aligned} d &= \mathcal{G}(\mathcal{X}_{true}) + \varsigma \\ &= d_{true} + \varsigma \end{aligned} \quad (1.186)$$

The forward problem involves finding d given \mathcal{X} . However, our focus lies in estimating \mathcal{X} given d , a task commonly referred to as the inverse problem. In numerous scenarios, we can express model parameters as an n -component vector denoted by $\vec{\mathcal{X}}$. Conversely, if the data is presented as a finite number of data points, it can be represented as an m -element vector denoted by \vec{d} . Such scenarios are termed discrete inverse problems or parameter estimation problems. Solving inverse problems is challenging as there could be multiple models that adequately fit the data. Nevertheless, once a solution is obtained, it must be evaluated for its physical plausibility, its ability to predict data, and its alignment with other constraints. We will now discuss a very well-known inverse problem, linear regression.

Linear regression Linear regression is the problem of finding a model, linear in its parameters, that approximately fits a set of data. Given a set of m observations \vec{d} and n model parameters represented as n -component vector $\vec{\mathcal{X}}$ which we want to estimate, we can express this discrete linear inverse problem as

$$\mathcal{G}\vec{\mathcal{X}} = \vec{d} \quad (1.187)$$

For simplicity, we assume that \mathcal{G} has full column rank ⁶. In many cases, the dimension of the column space of \mathcal{G} is smaller than $\vec{\mathcal{X}}$, and a noisy data \vec{d} will generally lie outside of the column space of \mathcal{G} . In these cases, there is no solution $\vec{\mathcal{X}}$ that satisfies Eq. (1.187) exactly. An approximate solution $\vec{\mathcal{X}}$ can be found such that it minimizes some measure of misfit between the \vec{d} and $\mathcal{G}\vec{\mathcal{X}}$. In this context, we define residual vector \vec{r}

$$\vec{r} = \mathcal{G}\vec{\mathcal{X}} - \vec{d} \quad (1.188)$$

⁶A m by n matrix \mathbf{M} has a full column rank if the rank of \mathbf{M} is equal to n , where the rank of the matrix \mathbf{M} is defined as the dimension of column space of \mathbf{M} .

One of the measures of misfit is the 2-norm of the residual vector, which is defined as follows:

$$\|\vec{r}\|_2 = \sqrt{\vec{r}^T \vec{r}}, \quad (1.189)$$

and the model that minimizes this 2-norm is called the least-squares solution. The least squares solution has the advantage that it is geometrically intuitive to understand, and it is statistically the most likely solution if the data errors are normally distributed. It can be obtained by projecting \vec{d} onto column space of \mathcal{G} .

$$\mathcal{G}\vec{x}_{ls} = \mathbf{p}_A \vec{d} \quad (1.190)$$

where $\mathbf{p}_A \vec{d}$ denotes the projection of \vec{d} onto column space of \mathcal{G} and subscript *ls* denotes least squares. Then $\mathcal{G}\vec{x}_{ls} - \vec{d}$, becomes perpendicular to the column space of \mathcal{G} that in turn implies that each of the columns of \mathcal{G} will be orthogonal to $\mathcal{G}\vec{x}_{ls} - \vec{d}$ which can be succinctly expressed as

$$\mathcal{G}^T(\mathcal{G}\vec{x}_{ls} - \vec{d}) = 0 \quad (1.191)$$

Rearranging the above equation leads to a system of equations known as normal equations:

$$\mathcal{G}^T \mathcal{G} \vec{x}_{ls} = \mathcal{G}^T \vec{d} \quad (1.192)$$

From the above equation, we find the least-squares solution to be

$$\vec{x}_{L2} = (\mathcal{G}^T \mathcal{G})^{-1} \mathcal{G}^T \vec{d} \quad (1.193)$$

It can be shown that if \mathcal{G} has a full column rank, then $(\mathcal{G}^T \mathcal{G})^{-1}$ exists.

L2-norm: A maximum likelihood solution Considering that the data is derived from measurements that inherently include random errors, it becomes essential to seek a statistically sound solution. One of these approaches is maximum likelihood estimation, which raises the question: given the observations, their statistical attributes, and a mathematical model for the forward problem, what model would be most likely to produce these observations?

Let's assign a probability density $p_i(d_i | \vec{\mathcal{X}})$ to the individual observations d_i , given a model $\vec{\mathcal{X}}$. Assuming the observations are independent, we can assign a joint probability density to all observations as follows:

$$p(\vec{d} | \vec{\mathcal{X}}) = \prod_{i=1}^m p_i(d_i | \vec{\mathcal{X}}) \quad (1.194)$$

The term $p(\vec{d} | \vec{\mathcal{X}})$ in the equation above is also known as the likelihood function and can be denoted as $\mathcal{L}(\vec{d} | \vec{\mathcal{X}}) \equiv p(\vec{d} | \vec{\mathcal{X}})$. It's also apparent that for models with an extremely low likelihood of producing the observed data, the likelihood value would approach zero. Conversely, models that are relatively more likely to generate the observed data would yield a higher likelihood value. This is what the maximum likelihood principle also proposes. It is interesting that for such discrete linear inverse problems, given the data errors are independent and normally distributed, the maximum likelihood solution is indeed the least-square solution. To see this, let's assume

that the data errors are independent and are distributed normally with zero mean and standard deviation σ_i for each observation d_i . The joint probability density or likelihood is given by

$$\mathcal{L}(\vec{d} \mid \vec{\mathcal{X}}) = \frac{1}{(2\pi)^{m/2} \prod_{i=1}^m \sigma_i} \prod_{i=1}^m e^{-\frac{1}{2}(d_i - (\mathcal{G}\vec{\mathcal{X}})_i)^2 / \sigma_i^2} \quad (1.195)$$

The maximization of the above equation can be considered as a minimization of the following equation:

$$\min \sum_{i=1}^m \frac{(d_i - (\mathcal{G}\vec{\mathcal{X}})_i)^2}{\sigma_i^2} \quad (1.196)$$

Defining $W = \text{diag}(1/\sigma_1, 1/\sigma_2, \dots, 1/\sigma_m)$, we can write down the weighted system of equations as the following:

$$\mathcal{G}_w \vec{\mathcal{X}} = \vec{d}_w \quad (1.197)$$

where $\mathcal{G}_w = W\mathcal{G}$ and $\vec{d}_w = W\vec{d}$

The least-squares solution corresponding to the Eq. (1.197) leads to

$$\vec{\mathcal{X}}_{l2} = (\mathcal{G}_w^T \mathcal{G}_w)^{-1} \mathcal{G}_w^T \vec{d}_w \quad (1.198)$$

Now, calculating the resulting residual vector $\vec{d}_w - \mathcal{G}_w \vec{\mathcal{X}}_{l2}$ and evaluating the corresponding square of the 2-norm leads to

$$\|\vec{d}_w - \mathcal{G}_w \vec{\mathcal{X}}_{l2}\|_2^2 = \sum_{i=1}^m \frac{(d_i - (\mathcal{G}\vec{\mathcal{X}}_{l2})_i)^2}{\sigma_i^2} \quad (1.199)$$

Therefore, the least-squares solution to $\mathcal{G}_w \vec{\mathcal{X}} = \vec{d}_w$ is indeed the maximum likelihood solution. Another important aspect of the least-squares solution is that it is unbiased. To see that let's calculate the expectation value of $\vec{\mathcal{X}}_{l2}$ solution:

$$E[\vec{\mathcal{X}}_{l2}] = (\mathcal{G}_w^T \mathcal{G}_w)^{-1} \mathcal{G}_w^T E[\vec{d}_w] \quad (1.200)$$

Since $E[\zeta] = 0$, it implies $E[\vec{d}_w] = \vec{d}_{true,w}$ and $\mathcal{G}_w \vec{\mathcal{X}}_{true} = \vec{d}_{true}$, it leads to the following:

$$\begin{aligned} \mathcal{G}_w^T \mathcal{G}_w \vec{\mathcal{X}}_{true} &= \mathcal{G}_w^T \vec{d}_{true,w} \\ &= \vec{\mathcal{X}}_{true} \end{aligned} \quad (1.201)$$

In the following sections, we will discuss the Bayesian analysis, which is an alternative approach to solving inverse problems.

1.2.2 Bayesian analysis

Bayesian analysis deals with conclusions about a parameter θ given some observed data y . In order to make some statements about θ given y , we need to first define a *joint probability density* for θ and y . It can be written as a product of two densities: a prior density $p(\theta)$ and the data distribution $p(y \mid \theta)$.

$$\begin{aligned} p(\theta, y) &= p(\theta) p(y \mid \theta) \\ &= p(y) p(\theta \mid y) \end{aligned} \quad (1.202)$$

where the second line of the above equation follows from the properties of conditional probability. Rearranging Eq. (1.202) yields the following equation:

$$p(\theta | y) = \frac{p(y | \theta) p(\theta)}{p(y)} \quad (1.203)$$

which is known as *Bayes' theorem* and $p(y) = \sum_{\theta} p(\theta) p(y | \theta)$, which is independent of θ and for a fixed y can be considered a constant. It is also known as the *model evidence*.

Before delving into Bayesian analysis, it's essential to briefly discuss the two fundamental distinctions between classical approaches, as outlined in the previous section, and the Bayesian approach. Firstly, in the classical approach, we seek to estimate a specific but unknown model denoted as m_{true} , whereas in the Bayesian approach, the model itself is not deterministic but rather a random variable. The solution is presented in the form of a probability distribution, allowing us to address various inquiries about the model. Secondly, the Bayesian approach inherently incorporates prior information about the solution, such as strict constraints or intuition derived from experience. This information can be mathematically expressed through the prior density/distribution for the model.

In principle, the posterior density $p(\theta | y)$ can be calculated using Bayes' theorem, assuming a given model that generates the data y . We can generate a grid on the parameter space θ , covering the important regions of the posterior density and calculating the right-hand side of the Eq. (1.203) would give us an estimate of posterior density $p(\theta)$ at those values of θ . For a higher dimensional parameter space, generating a dense multidimensional grid over the parameter space and evaluation of $p(\theta | y)$ would become prohibitively expensive. For instance, a model with 15 parameters would require a multidimensional grid over θ with N^{15} points in the parameter space, where N is the number of points in each dimension. Even for a modest value of $N = 10$, the number of points would be a whopping 10^{15} . The evaluation of $p(\theta | y)$ at each of those 10^{15} points would take ~ 32000 years, assuming each evaluation takes ~ 1 ms. Of course, 10 points in each dimension might not be dense enough to pick up the important features of the posterior density. This simple example highlights the importance of alternative strategies for estimating the posterior density $p(\theta | y)$. There exist many sampling methods to draw representative samples from the posterior density. We will now discuss some of these methods in the following sections.

1.2.3 Sampling

Direct approximation

Given the target density $p(\theta)$, evaluated at a set of evenly spaced values $\theta_1, \theta_2, \dots, \theta_N$, that cover a broad range of the parameter space for θ , one can draw samples from $p(\theta)$ using the inverse-CDF method, where CDF stands for the cumulative distribution function (CDF). The CDF of a one-dimensional distribution, $p(\theta)$, is defined by

$$F(\theta) = \begin{cases} \sum_{\theta^* < \theta} p(\theta^*) & \text{if } p \text{ is discrete.} \\ \int_{-\infty}^{\theta} p(\theta^*) d\theta^*, & \text{if } p \text{ is continuous.} \end{cases} \quad (1.204)$$

The inverse CDF can be used to obtain random samples from the distribution $p(\theta)$ as follows:

1. First, a random value, U is drawn from $\mathcal{U}(0, 1)$ which denotes a uniform distribution on $[0, 1]$.
2. Set $\theta = F^{-1}(U)$. The F may not necessarily be a one-to-one function, but $F^{-1}(U)$ is unique with probability 1. The value θ will be a random draw from $p(\theta)$ as long as $F^{-1}(U)$ is simple to calculate (see Theorem 2.1.10, p.54. [60]).

Let's take an example of exponential distribution, $p(\theta) = \lambda \exp(-\lambda \theta)$. We want to draw some samples from this distribution, which is evaluated at some evenly spaced θ values. We first estimate an empirical CDF using these θ values and then interpolate it using a cubic spline to get a smooth CDF function. Once we have the CDF interpolant, we can invert it, as mentioned earlier, to generate random samples from $p(\theta)$.

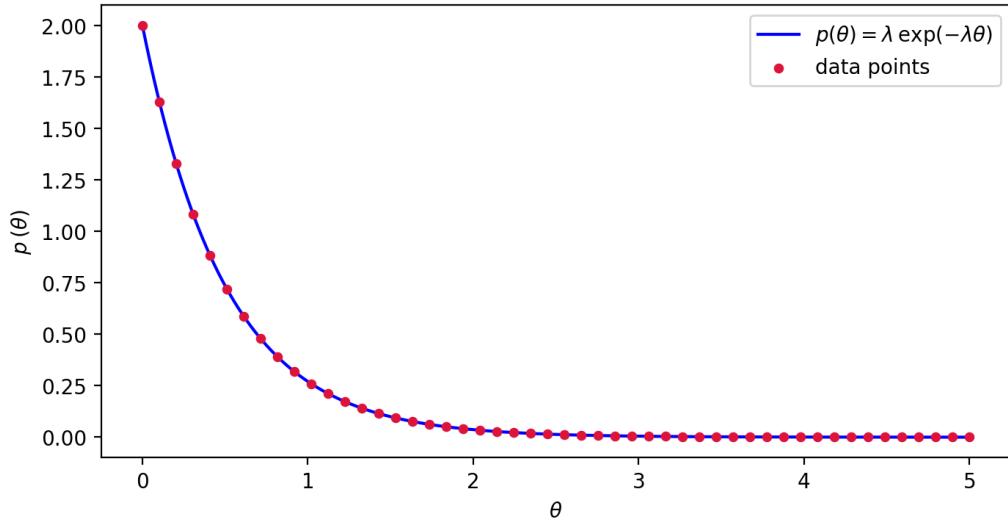


Figure 1.22: The red dots represent the values of θ where target density $p(\theta)$ is given. The task is to use these θ values and their corresponding densities $p(\theta)$ to draw random samples from the $p(\theta)$ itself.

As evident from Fig. 1.23, $N = 10$ data points are not enough for the inverse CDF method to accurately sample from the target density $p(\theta)$, while with $N = 50$ data points, the samples obtained from the inverse CDF method agree well with the true distribution. This illustration shows the limitation of the direct approximation method. If the points θ_i are spaced closely enough and mostly cover high-density areas, this method works well. Clearly, in high-dimensional multivariate problems, the number of points needed would become so large that it would become prohibitively expensive to compute $p(\theta)$ at every point of the multidimensional grid.

Rejection sampling

Rejection sampling [61] is another method to sample from a distribution when it might not be possible to draw samples from it directly. Suppose we want to draw a single random draw from a distribution $p(\theta)$. To perform rejection sampling, a positive function $q(\theta)$ for all θ , is required for which $p(\theta) > 0$ that has the following properties:

- A sample can be drawn from a probability distribution proportional to q .

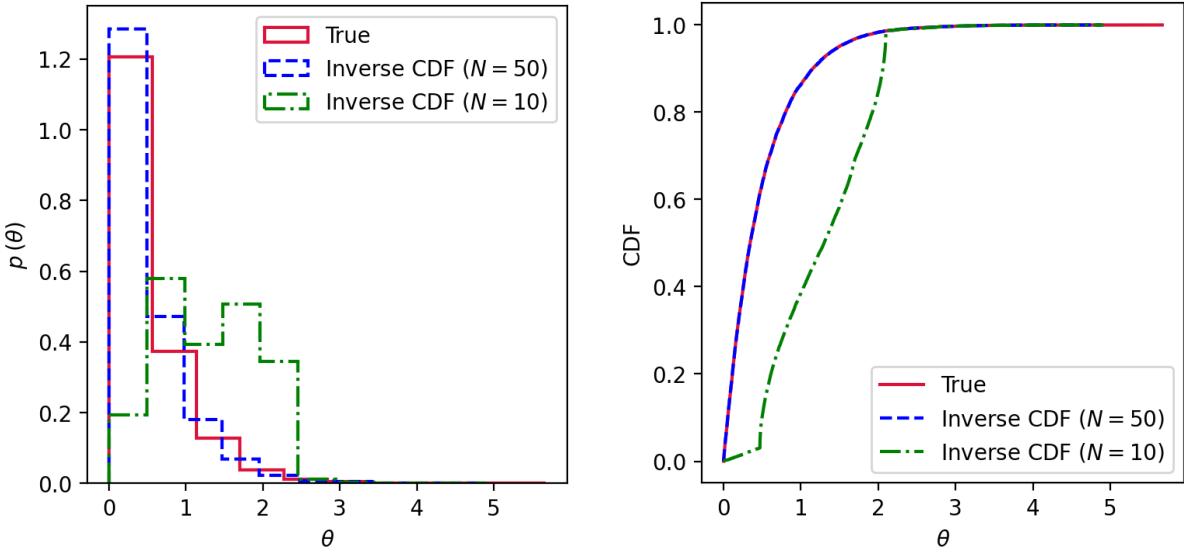


Figure 1.23: In the left panel, the red histogram shows samples directly drawn from the true distribution, while blue and green histograms of the samples are obtained using the inverse CDF method for $N = 50$ and $N = 10$ data points, respectively. On the right panel, the corresponding CDFs are plotted.

- There must exist some known constant M for which the *importance ratio* $p(\theta)/q(\theta) \leq M$ for all θ .

This algorithm proceeds in the following steps:

1. A random sample θ is drawn from the distribution proportional to $q(\theta)$.
2. θ is accepted as a draw from $p(\theta)$ with probability $p(\theta)/M q(\theta)$.

Fig. 1.24 shows the target distribution (an inverse-Gamma distribution) given by

$$p(\theta) = \frac{x^{-a-1}}{\Gamma(a)} \exp\left[-\frac{1}{x}\right] \quad (1.205)$$

whereas the proposal distribution $q(\theta)$ is a gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ with a given mean μ and standard deviation σ .

An effective option for the proposal distribution, $q(\theta)$, might involve a distribution that is approximately proportional to $p(\theta)$. Naturally, if we have q proportional to p with an appropriate M , each draw during step 2 would be accepted with probability 1. This algorithm possesses a self-monitoring feature, as only a few samples are accepted when the method is not operating efficiently.

Importance sampling

It is a method to estimate the expectation values using a random sample drawn from an approximate distribution to the target distribution. Let's consider a case where we want to compute $E(f(\theta))$ but cannot generate random samples from the target distribution, $p(\theta)$.

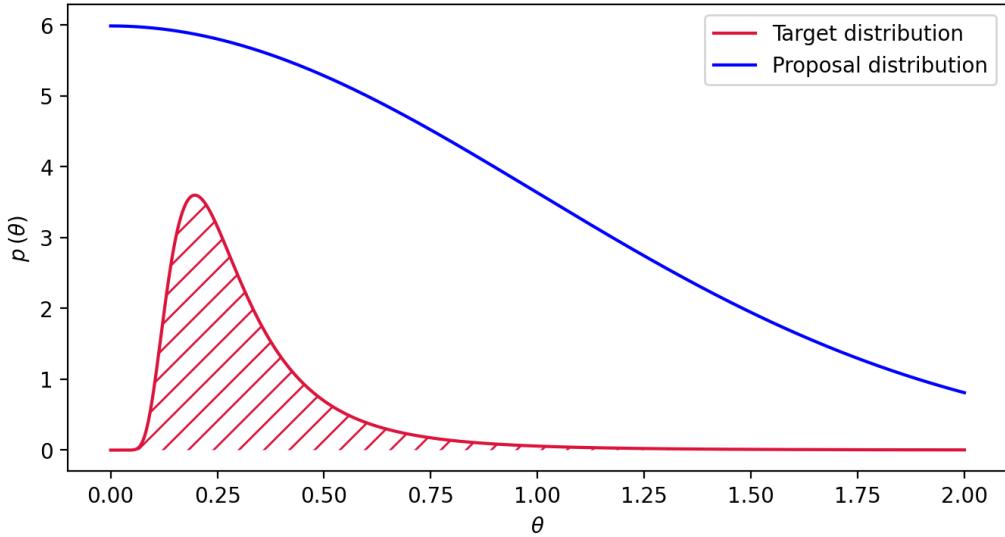


Figure 1.24: The red curve represents the target distribution $p(\theta)$, which is an inverse-Gamma distribution. The blue curve is the proposal distribution $q(\theta)$, which in this case is a Gaussian distribution. The probability that a sample θ is accepted is the ratio of the height of the lower curve to the height of the higher curve at the value θ .

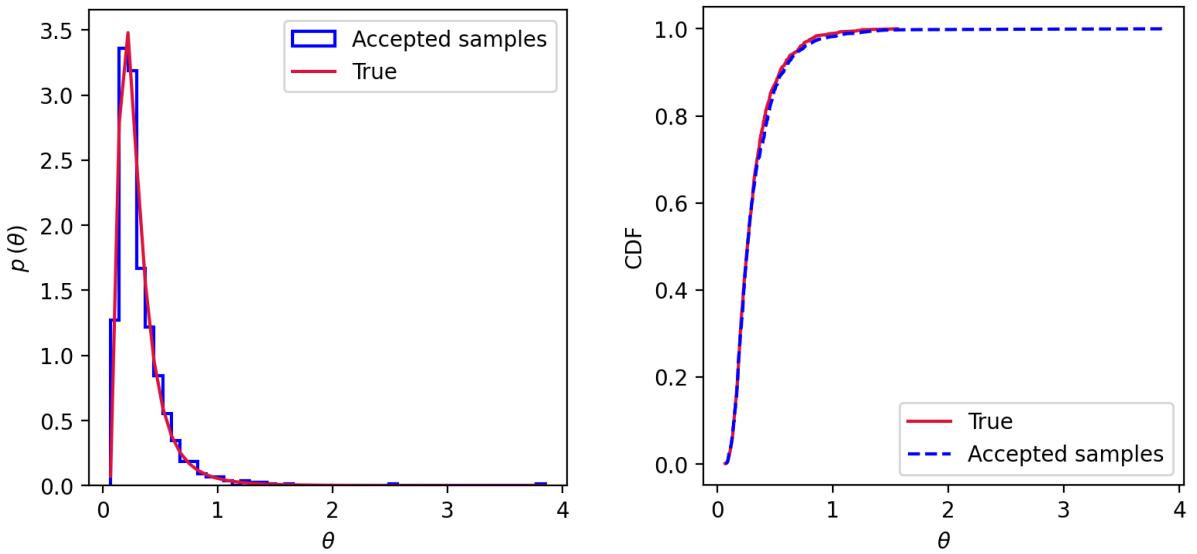


Figure 1.25: In the left panel, the red curve is the true distribution $p(\theta)$, while the blue histogram shows the accepted samples in the rejection sampling algorithm. On the right panel, their corresponding CDFs are shown.

If $q(\theta)$ is the distribution from which we can draw random samples, then we can estimate $\mathbb{E}(f(\theta))$ as follows:

$$\begin{aligned}\mathbb{E}(f(\theta)) &= \frac{\int f(\theta) p(\theta) d\theta}{\int p(\theta) d\theta} \\ &= \frac{\int [f(\theta) p(\theta) / q(\theta)] q(\theta) d\theta}{\int [p(\theta) / q(\theta)] q(\theta) d\theta}.\end{aligned}\tag{1.206}$$

which can be estimated using N number of draws $\theta_1, \dots, \theta_N$ from $q(\theta)$ by the expression,

$$\frac{\frac{1}{N} \sum_{n=1}^N f(\theta_n) w(\theta_n)}{\frac{1}{N} \sum_{n=1}^N w(\theta_n)} \quad (1.207)$$

where $w(\theta_n) = p(\theta_n)/g(\theta_n)$ are called *importance ratios* or *importance weights*. Note that $p(\theta)$ may not necessarily be normalized.

For an illustration of this method (see the left panel in Fig. 1.26), suppose we want to estimate $\mathbb{E}(f(\theta))$, where $f(\theta) = \frac{1}{1 + \exp(-\theta)}$ and our target distribution is an exponential distribution. Let's choose the sampling distribution $q(\theta)$ to be a truncated normal distribution (so that all the samples are positive) with $\mu = 0$ and $\sigma = 1$.

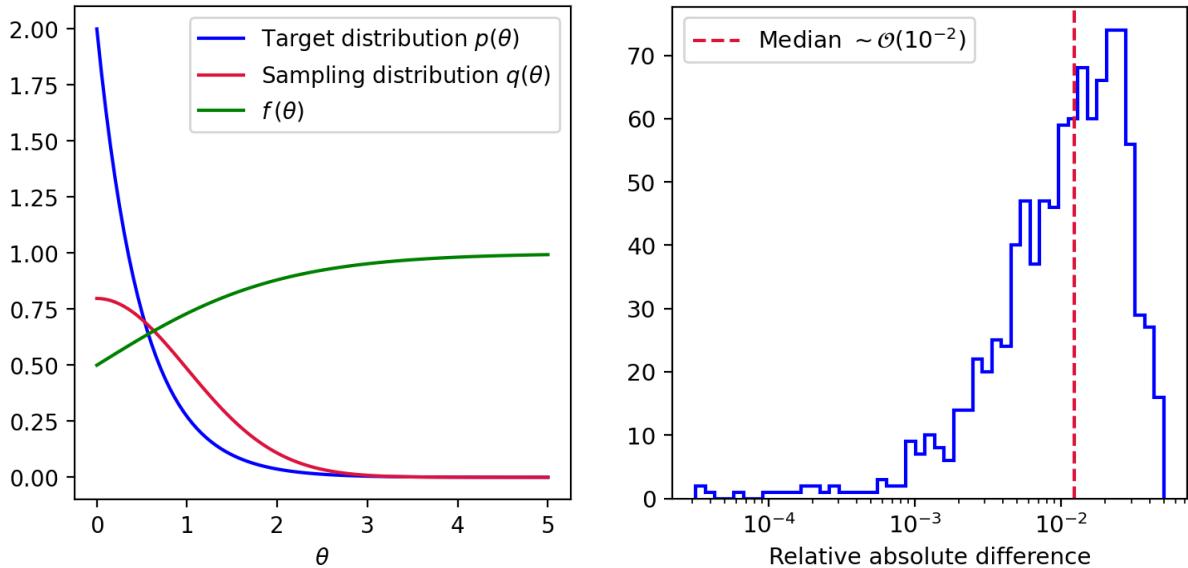


Figure 1.26: The left panel shows the target (blue) and sampling distribution (red), and the function $f(\theta)$ (green) whose expectation value is to be estimated. The right panel shows a histogram of the relative absolute difference between $\mathbb{E}(f(\theta))_{p(\theta)}$ and $\mathbb{E}(f(\theta))_{q(\theta)}$. Evidently, the relative difference is mostly $\lesssim 5\%$.

From Fig. 1.26, we can conclude the estimation of the integral in Eq. (1.206) can be estimated fairly accurately for a suitable choice of the sampling distribution $q(\theta)$. Moreover, a fairly precise estimate of the integral can be obtained if we choose $q(\theta)$ such that $\frac{fp}{q}$ is nearly a constant. One of the limitations of importance sampling [62] is that it may not be useful when the importance ratios vary substantially, particularly when they are small with a high probability but with a low probability are huge. It usually happens if fp has wide tails compared to q as a function of θ .

Markov chain (MC) simulation

A Markov chain [63] is defined by a sequence of random elements $\theta_1, \theta_2, \dots$ if the conditional distribution of θ_{n+1} given $\theta_1, \theta_2, \dots, \theta_n$ depends only on θ_n . It has a *stationary transition distribution* if the conditional distribution of θ_{n+1} given θ_n does not depend on n . A Markov chain simulation is typically used when it is not possible or not computationally efficient to

directly sample θ from the target density $p(\theta | y)$. Instead, samples are drawn *iteratively* from a distribution that becomes closer to $p(\theta | y)$ as simulation proceeds. The key idea is to construct a Markov chain whose stationary distribution is target distribution $p(\theta)$, and the simulation runs for long enough to reach a point where the stationary distribution is very close to the target density. In other words, in each iteration, we are sampling from a distribution that is becoming close to the target distribution. An MC simulation starts from some point θ_0 and then in each iteration, a new point θ_i is proposed from a *transition/proposal distribution* $T_i(\theta_i | \theta_{i-1})$, where i denotes the iteration number. The proposal distribution is constructed in such a way that the MC converges to the target distribution $p(\theta)$. In the following sections, we will discuss some basic MC simulation methods such as Gibbs sampling, Metropolis, and Metropolis-Hastings algorithms.

Gibbs sampling Gibbs sampling [64] is a kind of MC algorithm that is found to be useful in several multidimensional problems. It is also called alternating conditional sampling because it involves drawing samples for one of the model parameters from a distribution conditioned on the rest of the model parameters. In other words, the parameter space is divided into d components or subvectors, $\theta = (\theta_1, \theta_2, \dots, \theta_d)$ and in every iteration t the sample for each parameter θ_j is drawn from a distribution $p(\theta_j | \theta_{-j}^{t-1}, y)$, where θ_{-j}^{t-1} represents all the components of θ , except for θ_j , at their current values:

$$\theta_{-j}^{t-1} = (\theta_1^t, \dots, \theta_{j-1}^t, \theta_{j+1}^{t-1}, \dots, \theta_d^{t-1}). \quad (1.208)$$

Thus, each subvector θ_j is updated conditional on the latest values of the other components of θ , which are iteration t values for the components already updated and the iteration $t - 1$ values for the others. Let's try to understand the Gibbs sampling algorithm with the help of a simple example. Consider a single observation $\vec{y} = (y_1, y_2)$ drawn from a bivariate normal distribution with an unspecified mean $\vec{\theta} = (\theta_1, \theta_2)$ and covariance variance matrix defined as $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. Assuming a uniform prior distribution for the parameter θ , the resulting posterior distribution becomes

$$p(\vec{\theta} | \vec{y}) \sim \mathcal{N}(\vec{y}, \Sigma) \quad (1.209)$$

Using the properties of multivariate normal distribution, we can write the conditional posterior distributions for each parameter as follows:

$$\begin{aligned} p(\theta_1 | \theta_2, y) &\sim \mathcal{N}(y_1 + \rho(\theta_2 - y_2), (1 - \rho^2)) \\ p(\theta_2 | \theta_1, y) &\sim \mathcal{N}(y_2 + \rho(\theta_1 - y_1), (1 - \rho^2)) \end{aligned} \quad (1.210)$$

Gibbs sampling proceeds in the following steps:

1. A random sample θ_1 is drawn from $p(\theta_1 | \theta_2^{\text{init}}, y)$ where θ_2^{init} is the some initial value of θ_2 .
2. Now a sample θ_2 is drawn from $p(\theta_2 | \theta_1, y)$ where θ_1 is the value drawn in the first step.
3. Then put $\theta_2 = \theta_2^{\text{init}}$, and then all the steps repeat till the iterations end.

For this example, we choose $\rho = -0.6$ and generate four independent sequences which start from $(\pm 5.5, \pm 5.5)$.

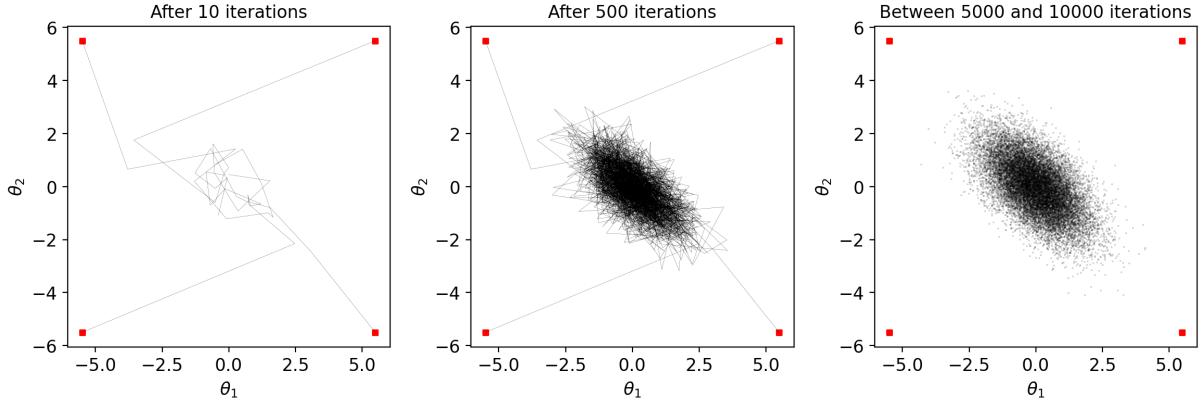


Figure 1.27: Gibbs sampling for four independent sequences starting from overdispersed locations in the parameter space, represented by solid red squares, for a bivariate normal distribution with correlation $\rho = -0.6$. The first panel shows the component-wise updating of Gibbs iterations. The second panel shows the distribution of θ_1 and θ_2 after 500 iterations, which have reached close to convergence. The last panel displays the latter half of the sequences, which have now converged, and the obtained samples accurately represent the target distribution (example reproduced from [65]).

Metropolis algorithm and Metropolis-Hastings algorithm Metropolis algorithms [66, 67] belong to a family of MC simulations used for generating samples from Bayesian posterior distributions. We will now discuss a brief overview of the basic Metropolis algorithm, followed by its extension to the Metropolis-Hastings algorithm.

The Metropolis algorithm The Metropolis algorithm uses a random walk with acceptance/rejection criteria to converge to the target distribution. It proceeds as described in the following steps:

1. Choose a starting point θ_0 . It can also be drawn from a starting distribution $p_0(\theta)$ such that $p(\theta_0 | y) > 0$.
2. For $t = 1, 2, \dots$, where t indicates the iteration number.

- Draw a sample θ^* from a proposal distribution at iteration t , $J_t(\theta^* | \theta^{t-1})$. The proposal distribution must be *symmetric*, satisfying the condition $J_t(\theta_a | \theta_b) = J_t(\theta_b | \theta_a)$ for all θ_a, θ_b , and iterations.
- Estimate the ratio of densities,

$$r = \frac{p(\theta^* | y)}{p(\theta^{t-1} | y)}. \quad (1.211)$$

- Set

$$\theta^t = \begin{cases} \theta^*, & \text{with probability } \min(r, 1). \\ \theta^{t-1}, & \text{otherwise.} \end{cases} \quad (1.212)$$

The above step can be performed by generating a uniform random number.

To demonstrate the Metropolis algorithm, we again use a simple example of bivariate normal distribution where the target distribution is a bivariate normal distribution, $p(\theta | y) = \mathcal{N}(\theta | 0, \Sigma)$, where $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ with $\rho = 0.9$. The proposal distribution is also chosen to be a bivariate normal distribution centered on the current iteration with a jumping scale equal to 1/4 the size: $J_t(\theta^* | \theta^{t-1}) \sim \mathcal{N}(\theta^* | \theta^{t-1}, 0.25^2 \Sigma)$, which is also symmetric.

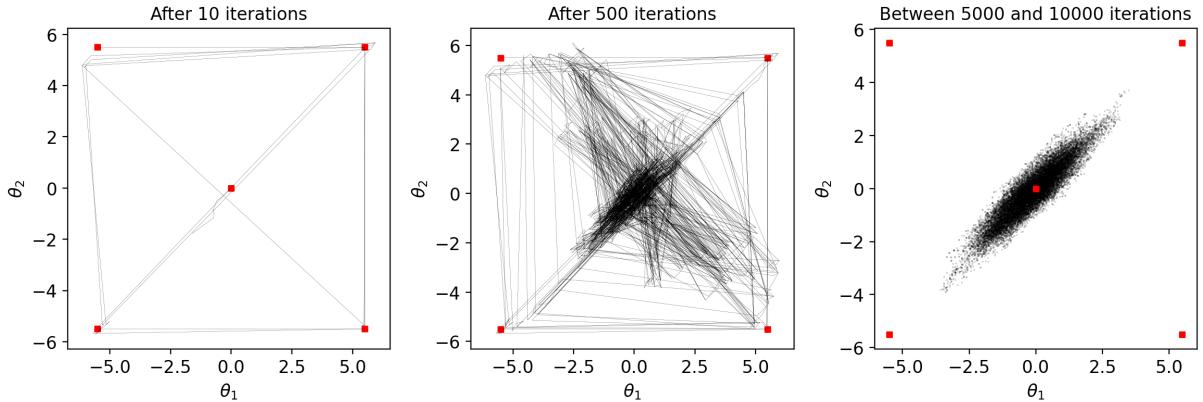


Figure 1.28: Metropolis algorithm for five independent sequences starting from overdispersed locations in the parameter space, represented by solid red squares, for a bivariate normal distribution with correlation $\rho = 0.9$. The first panel shows the sampled distribution from the first 10 iterations, which is far from convergence. The second panel shows the distribution after 500 iterations and looks close to convergence. In the last panel, samples from the latter half of the sequences are shown, which have now converged and represent the highly correlated draws from the target distribution (example reproduced from [65]).

The Metropolis-Hastings algorithm The Metropolis algorithm, as described in the last section, can be generalized to the Metropolis-Hastings algorithm by relaxing the requirement of symmetric proposal distributions. In other words, $J_t(\theta_a | \theta_b)$ may not necessarily be equal to $J_t(\theta_b | \theta_a)$. Asymmetric proposal distributions can be useful in speeding up the random walk. Additionally, the r in Eq. (1.211) is replaced by a ratio of ratios:

$$r = \frac{p(\theta^* | y)/J_t(\theta^* | \theta^{t-1})}{p(\theta^{t-1} | y)/J_t(\theta^{t-1} | \theta^*)}. \quad (1.213)$$

Let's see how the Metropolis algorithm, indeed, converges to the target distribution. To begin, the generated sequences should constitute a Markov Chain (MC) with a unique stationary distribution. This is fulfilled when the MC meets the criteria of irreducibility, aperiodicity, and non-transience. With the exception of some straightforward instances, these latter conditions are satisfied for most random walks on appropriate distributions. Furthermore, the condition of irreducibility necessitates that the proposal distribution J_t can transition to all states with a non-zero probability. The Metropolis algorithm guarantees that the stationary distribution associated with MC is the target distribution. In order to prove it, let's consider that the

algorithm starts at an iteration $t - 1$ where a sample is drawn θ^{t-1} from the target distribution $p(\theta | y)$. Suppose two points θ_a and θ_b , are drawn from $p(\theta | y)$ and labelled in such a way that $p(\theta_a | y) \leq p(\theta_b | y)$. The corresponding transition probability from θ_a to θ_b is given by

$$p(\theta^{t-1} = \theta_a, \theta^t = \theta_b) = p(\theta_a | y) J_t(\theta_b | \theta_a), \quad (1.214)$$

where the probability of acceptance is 1 by construction. Now, the transition probability from θ_b to θ_a ,

$$\begin{aligned} p(\theta^t = \theta_a, \theta^{t-1} = \theta_b) &= p(\theta_b | y) J_t(\theta_a | \theta_b) \left(\frac{p(\theta_a | y)}{p(\theta_b | y)} \right) \\ &= p(\theta_a | y) J_t(\theta_a | \theta_b), \end{aligned} \quad (1.215)$$

which is the same as Eq. (1.214) since we assumed a symmetric proposal distribution. Following similar arguments, we can prove the validity of the Metropolis-Hastings algorithm.

Nested sampling

Markov Chain Monte Carlo (MCMC) techniques [68, 69, 70], while effective in generating samples for a diverse range of problems, can encounter challenges when dealing with posteriors characterized by widely separated modes. Moreover, due to its generation of samples in proportion to the posterior, calculating the model evidence (refer to Eq. (1.203)) using these samples becomes challenging. The Nested sampling [71] is a novel approach to estimate both the posterior and evidence. This method addresses the limitations of MCMC by partitioning the posterior into numerous nested slices, generating samples from each slice, and then combining these samples with appropriate weights to reconstruct the target (posterior) distribution. Let us understand how Nested sampling is able to estimate evidence and the corresponding posterior as a byproduct of evidence estimation.

In contrast to MCMC methods, which primarily target direct sampling from the posterior, Nested sampling is designed for estimating the evidence, which is defined as

$$\begin{aligned} \mathcal{Z} &\equiv \int_{\Omega_{\vec{\theta}}} p(\vec{\theta}) d\vec{\theta} \\ &= \int_{\Omega_{\vec{\theta}}} \mathcal{L}(\vec{\theta}) \pi(\vec{\theta}) d\vec{\theta}, \end{aligned} \quad (1.216)$$

and refactor the integral above to encompass the prior volume denoted as X within the enclosed parameter space.

$$\mathcal{Z} = \int_{\Omega_{\vec{\theta}}} \mathcal{L}(\vec{\theta}) \pi(\vec{\theta}) d\vec{\theta} = \int_0^1 \mathcal{L}(X) dX \quad (1.217)$$

where $\mathcal{L}(X)$ corresponds to an equal-likelihood contour which defines the edge of the volume X , whereas the prior volume is defined as the fraction of the prior with the likelihood $\mathcal{L}(\vec{\theta}) \geq \lambda$ and can be written as

$$X(\lambda) \equiv \int_{\Omega_{\vec{\theta}} : \mathcal{L}(\vec{\theta}) \geq \lambda} \pi(\vec{\theta}) d\vec{\theta} \quad (1.218)$$

A normalized prior implies that $X(\lambda = 0) = 1$ and $X(\lambda = \infty) = 0$, defining the bounds of integration for Eq. (1.217).

Generating samples Nested sampling, instead of sampling directly from the posterior, samples from the prior $\pi(\vec{\theta})$ subjected to a hard likelihood constraint λ , which might be difficult for an arbitrary prior $\pi(\theta)$ because of the drastic variations of the density across the parameter space. Nonetheless, the process can be simplified when the prior follows a standard uniform distribution (i.e., flat between 0 and 1) across all dimensions, resulting in a constant density within the interior of λ . This can be achieved employing a suitable “prior transform” function \mathcal{T} that maps a set of parameters $\vec{\Theta}$, characterized by a uniform prior distribution over the D -dimensional unit cube, to the desired parameter $\vec{\theta}$. It simplifies our task from directly sampling from the posterior $p(\vec{\theta})$ to a much easier task of repeatedly sampling uniformly within the transformed, constrained prior.

$$\pi'_\lambda(\vec{\Theta}) \equiv \begin{cases} 1/X(\lambda) & \mathcal{L}(\vec{\theta} = \mathcal{T}(\vec{\Theta})) \geq \lambda \\ 0 & \text{otherwise} \end{cases} \quad (1.219)$$

Since the exact value of $X(\lambda)$ at a given likelihood level $\lambda = \mathcal{L}(\vec{\theta})$ is unknown, we can construct an estimator \hat{X} with a known statistical distribution. It is evident from Eq. (1.219), $X(\mathcal{L})$ defines a CDF over \mathcal{L} . Therefore, we can define the corresponding probability distribution function (PDF) for \mathcal{L} as

$$P(\mathcal{L}) \equiv \frac{dX(\mathcal{L})}{d\mathcal{L}} = \frac{d}{d\mathcal{L}} \int_{\Omega_{\vec{\theta}}} \pi_{\mathcal{L}}(\vec{\theta}) d\vec{\theta} \quad (1.220)$$

Assuming it is possible to sample \mathcal{L} from its PDF $P(\mathcal{L})$, we can use the Probability Integral Transform (PIT)⁷ to sample $X(\mathcal{L})$ from a standard uniform distribution $\mathcal{U}(0, 1)$:

$$\mathcal{L}' \sim P(\mathcal{L}) \Rightarrow X(\mathcal{L}') \sim \mathcal{U}(0, 1) \quad (1.221)$$

where $\mathcal{U}(0, 1)$ is the standard Uniform distribution. It can also be extended for cases when the sampling is performed with a given threshold.

$$\mathcal{L}' \sim P(\mathcal{L} | \mathcal{L} > \lambda) \Rightarrow \frac{X(\mathcal{L}')}{X(\lambda)} \sim \mathcal{U}(0, 1) \quad (1.222)$$

which implies we directly sample from $\pi_\lambda(\vec{\theta})$ (bypassing λ and $P(\mathcal{L})$) to satisfy PIT:

$$\vec{\theta}' \sim \pi_\lambda(\vec{\theta}) \Rightarrow \frac{X(\mathcal{L}(\vec{\theta}'))}{X(\lambda)} \sim \mathcal{U}(0, 1) \quad (1.223)$$

Now, for a given prior volume X_{i-1} corresponding to a given likelihood level $\lambda_{i-1} = \mathcal{L}(\vec{\theta}_{i-1})$ after $i - 1$ iterations of this procedure, the current prior volume X_i will be

$$\hat{X}_i = U_i \hat{X}_{i-1} = \prod_{j=1}^i U_j \quad (1.224)$$

⁷The PIT refers to a theorem stating that if a random variable X follows a continuous distribution with a CDF given denoted by F_X , then a random variable defined as $Y = F_X(X)$ has a standard uniform distribution (see Theorem 2.1.10, p.54. [60]).

where $U_1, U_2, \dots, U_i \stackrel{\text{iid}}{\sim} \mathcal{U}(0, 1)$ are independent and identically distributed (iid) random variables drawn from the $\mathcal{U}(0, 1)$. Now, in order to decide whether to terminate the iterations, we need to estimate \hat{X}_i . The expectation value of \hat{X}_i could be a good choice:

$$\mathbb{E}[\hat{X}_i] = \mathbb{E}\left[\prod_{j=1}^i U_j\right] = \prod_{j=1}^i \mathbb{E}[U_j] = \left(\frac{1}{2}\right)^i \quad (1.225)$$

Similarly, we can also evaluate the expectation value of $\ln \hat{X}_i$ (also known as the geometric mean),

$$\ln \mathbb{E}[\hat{X}_i] = \sum_{j=1}^i \mathbb{E}[\ln U_j] \sim - \sum_{j=1}^i [E_j] = -i \quad (1.226)$$

where we have used the fact that

$U \sim \mathcal{U}(0, 1) \Rightarrow -\ln U \sim \text{Expo}$, where Expo is the standard exponential distribution and

$$E_1, E_2, \dots, E_i \stackrel{\text{iid}}{\sim} \text{Expo} \quad (1.227)$$

So far, we have considered sampling using only one point (also known as a live point), which evolves at each iteration in accordance with the constrained prior. We can either perform multiple nested sampling runs with a single live point or simply perform a single nested sampling run with many live points. In fact, as shown in [72, 73, 74], sampling with K live points is equivalent to combining K sets of independent samples derived using one live point each. We will now calculate the estimate \hat{X}_i of the prior volume in case of K live points. At any given iteration i , the current set of prior volumes $\{X_i^{[1]}, X_i^{[2]}, \dots, X_i^{[K]}\}$ corresponding to the K live points are uniformly distributed within the prior volume from the previous iteration X_{i-1} so that

$$X_i^{[j]} = U^{[j]} X_{i-1} \quad (1.228)$$

where $U^{[1]}, U^{[2]}, \dots, U^{[K]} \stackrel{\text{iid}}{\sim} \mathcal{U}(0, 1)$. Now we replace the live point with the lowest likelihood \mathcal{L}_i^{\min} (also called “dead point”) corresponding to the largest prior volume for which make an ordered list of prior volumes

$$X_i^{(j)} = U^{(j)} X_{i-1} \quad (1.229)$$

where (j) now indicates the position in the ordered list (from smallest to largest) and

$$U^{(j)} \sim \beta(j, K+1-j) \quad (1.230)$$

where $\beta(., .)$ is a beta distribution. Therefore, we can write an estimate of the prior volume based on K live points at iteration i as

$$\hat{X}_i = \prod_{j=1}^i U_j^{(K)} \quad (1.231)$$

where $U_i^{(K)}, U_i^{(K)}, \dots, U_i^{(K)}$ are the iid draws of the K^{th} standard uniform order statistic with marginal distribution $\beta(K, 1)$. Its arithmetic and geometric mean are given as follows:

$$\begin{aligned}\mathbb{E}(\hat{X}_i) &= \left(\frac{K}{K+1}\right)^i \\ \mathbb{E}(\ln \hat{X}_i) &= -\frac{i}{K}\end{aligned}\tag{1.232}$$

From the above equations, we can compute an estimate of the change in the prior volume at a

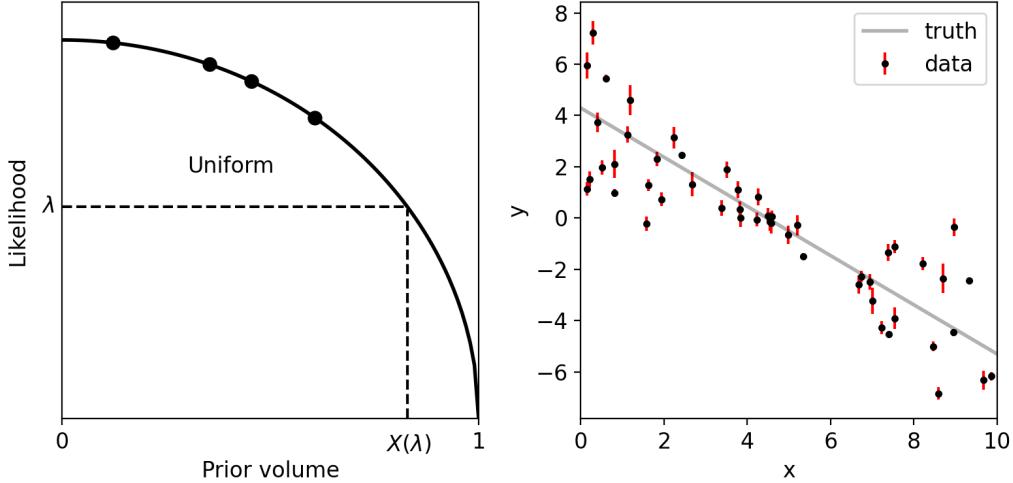


Figure 1.29: The figure in the left panel shows how 4 live points are sampled uniformly in $X < X(\lambda)$ (or in $\mathcal{L} < \mathcal{L}^*$). In the right panel, the linear model generating the observations is shown. The data also contains noise following a normal distribution. The gray line represents the true model, whereas the black dots (with error bars) denote the data. We want to estimate the model parameters using nested sampling. (left figure reproduced from [75])

given iteration i as

$$\mathbb{E}[\Delta \ln \hat{X}_i] = \mathbb{E}[\ln \hat{X}_i - \ln \hat{X}_{i-1}] = -\frac{1}{K}\tag{1.233}$$

This is also known as *exponential shrinkage* of the prior volume.

Termination Condition Since nested sampling aims to estimate the evidence, a natural stopping criterion is when the current set of live points covers almost the entire posterior distribution. In other words, we can stop sampling at a given iteration if

$$\Delta \ln \hat{\mathcal{Z}}_i \equiv \ln(\hat{\mathcal{Z}}_i + \Delta \hat{\mathcal{Z}}_i) - \ln(\hat{\mathcal{Z}}_i) < \epsilon\tag{1.234}$$

where $\Delta \hat{\mathcal{Z}}_i$ is the estimated remaining evidence we have yet to integrate over, and ϵ determines the tolerance. In principle, we can construct a strict upper bound on the remaining evidence $\Delta \hat{\mathcal{Z}}_i \lesssim \mathcal{L}^{\max} X_i$, where \mathcal{L}^{\max} is the maximum-likelihood value across the sample space and X_i is the prior volume at the current iteration. Since both \mathcal{L}^{\max} and X_i are not known exactly, we can work with their corresponding best estimates available. \mathcal{L}^{\max} across the sample space can be replaced by maximum likelihood among the live points at iteration i and \hat{X}_i is the estimated prior volume.

Evidence estimation The evidence integral can be approximated using a set of N dead points via

$$\mathcal{Z} = \int_0^1 \mathcal{L}(X) dX \approx \hat{\mathcal{Z}} = \sum_{i=1}^N \mathcal{L}_i(X_{i-1} - X_i) = \sum_{i=1}^N \hat{w}_i \quad (1.235)$$

where \hat{w}_i is estimated weight of each dead point.

Posterior estimation As a byproduct of evidence estimation, we can estimate the posteriors from the same set of N dead points by assigning each sample its associated *importance* weight as follows:

$$p(\vec{\theta}_i) = p(X_i) \equiv p_i \approx \hat{p}_i = \frac{\hat{w}_i}{\sum_{i=1}^N \hat{w}_i} = \frac{\hat{w}_i}{\hat{\mathcal{Z}}} \quad (1.236)$$

We use a simple example of linear regression to estimate the model parameters using nested sampling. The data-generating model is described by $y = mx + b$, where m and b are the parameters to be estimated. The noise in the data is modeled by a Gaussian distribution with variance underestimated by a fraction amount f , which we will also estimate along with other model parameters. The corresponding likelihood function is given by

$$\ln \mathcal{L}(y | x, \sigma, m, b, f) = -\frac{1}{2} \sum_n \left[\frac{(y_n - mx_n - b)^2}{s_n} + \ln(2\pi_n^2) \right] \quad (1.237)$$

where $s_n^2 = \sigma_n^2 + f^2(mx_n + b)^2$. We choose uniform priors over m, b , and $\ln f$. It can be represented in the following form:

$$p(m) = \begin{cases} 1/(m_{max} - m_{min}), & \text{if } m_{min} < m < m_{max} \\ 0, & \text{otherwise.} \end{cases} \quad (1.238)$$

Likewise, we can write down the uniform prior distribution over other parameters. These types of priors are often referred to as 'uninformative' because they reflect a lack of prior knowledge before performing the PE. Nevertheless, it is important to ensure that these priors encompass the significant regions of the posterior distribution. In scientific applications like GW data analysis, these priors are frequently inspired by astrophysical considerations. For example, all known neutron stars fall within the mass range of $1 M_\odot$ to $2.35 M_\odot$. Consequently, when analyzing binary neutron star systems, our priors must encompass this range. We used `dynesty` [72, 73], a Python implementation of a nested sampling algorithm for running the PE. Fig. 1.30 shows the posterior distributions over m, b and $\ln f$. The true value (as shown by the solid red line) lies within the 90% credible intervals (CI) of the posterior distributions.

Effect of priors and likelihoods on posterior

An essential aspect concerning the priors and likelihood is the potential conflict arising from their concentration in different regions of the parameter space. In such scenarios, the shape of the distribution tails significantly impacts the resulting posterior distribution. To demonstrate it, let's consider four cases (models) to fit one observation at $d = 0$ (taken from problem 13M6 on page 432 [76]). As evident from Fig. 1.31, the priors have a dramatic effect on the shape of

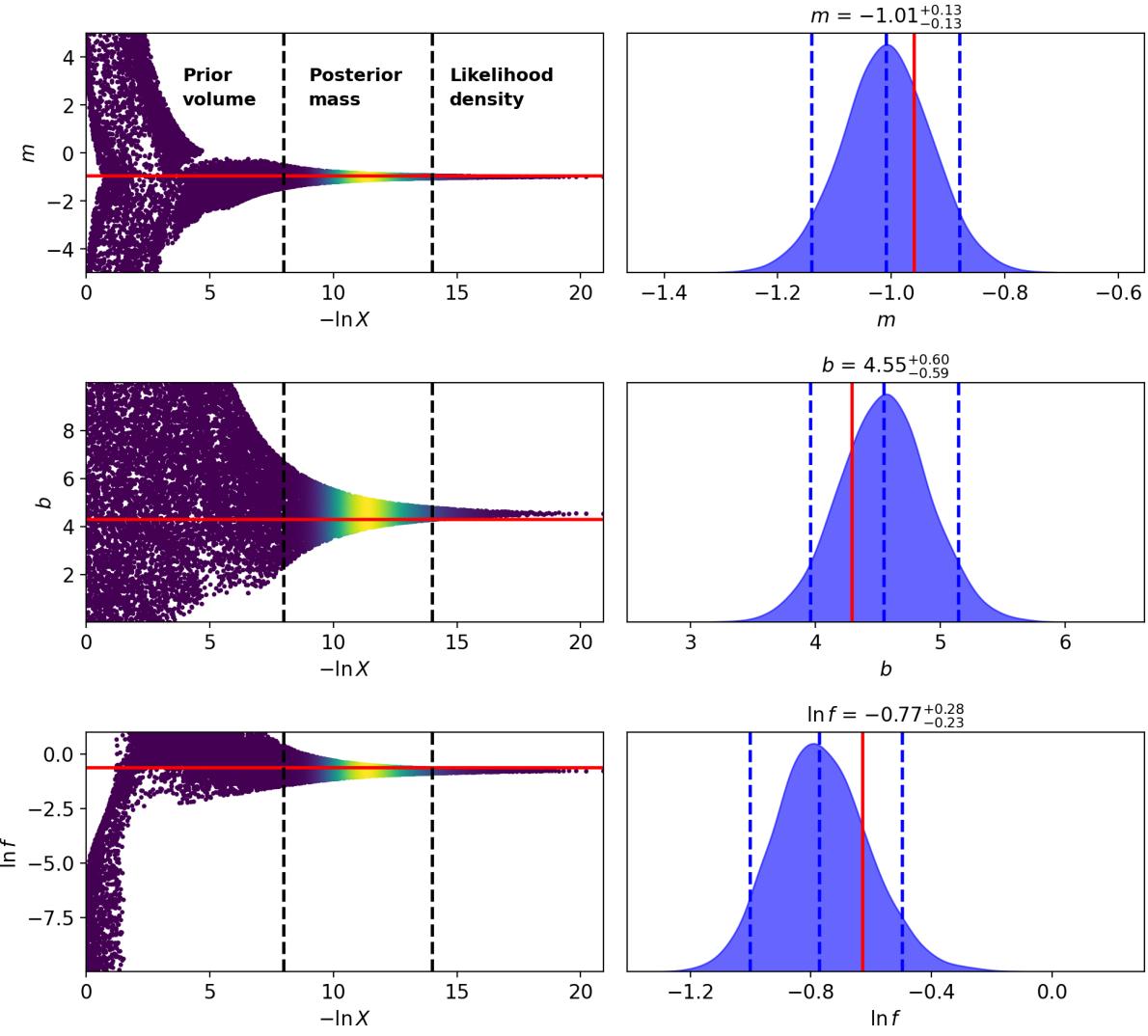


Figure 1.30: PE results using nested sampling are shown. The first column in the figure represents trace plots of the “dead” points positions as a function of $\ln X$, colored by the importance weight. The red solid line depicts the true value of the parameters. The PDF of the importance weight can be roughly categorized into distinct regions based on the influence of prior volume, posterior mass, and likelihood density. The posterior mass holds the highest relevance in posterior estimation, while evidence estimation also relies on the prior volume. The posterior distributions over m , b , and $\ln f$ are presented in the second column of the figure, with each posterior’s title indicating the median values and their respective 90% CIs. The true values are illustrated using a solid blue line.

the posterior. The Gaussian distribution is characterized by narrow tails, while the Student-t distribution exhibits heavier tails. When either the prior or the likelihood has heavier tails, the resulting posterior is skewed towards the side with heavier tails (as seen in case 2 and case 3). Conversely, when both the prior and the likelihood have narrow tails (case 1), the posterior represents a compromise between the two. In the case where both distributions have heavier tails (case 4), each distribution views the other mode as more plausible, resulting in a compromise in

the posterior, albeit not the optimal one.

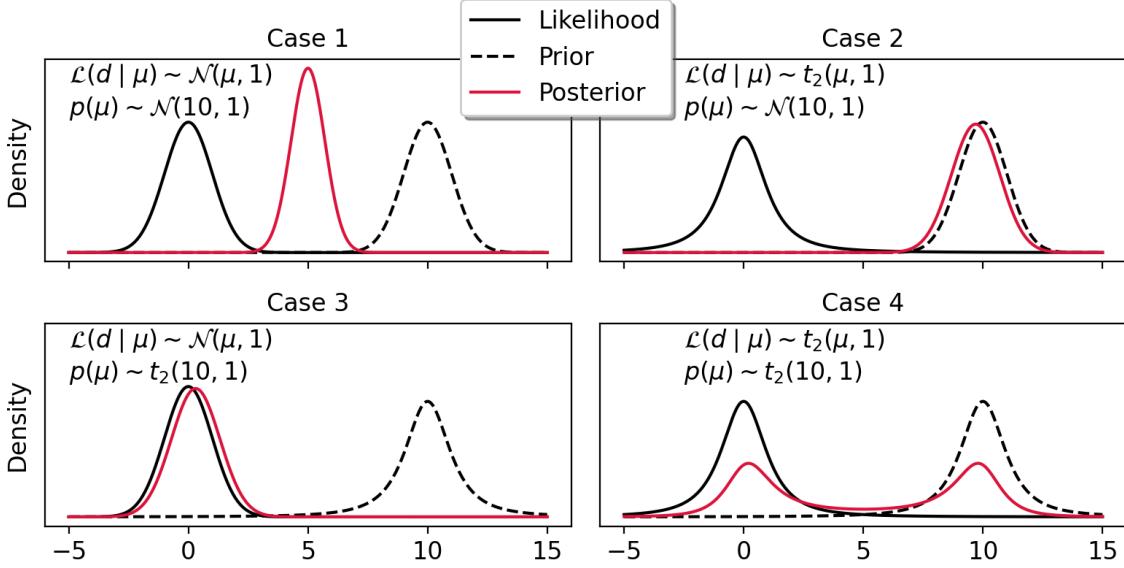


Figure 1.31: The effect of priors on the shape of the posterior distributions is shown.

In the next chapter, we discuss a fast PE method based on a combination of numerical linear algebra and meshfree approximation to reconstruct the compact binary sources.

1.3 Appendix

The Christoffel symbol is given by

$$\Gamma_{\mu\nu}^\rho = \frac{1}{2}g^{\partial\sigma}(\rho_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}) \quad (1.239)$$

The Riemann tensor is defined as

$$R^\mu_{\nu\rho\sigma} = \partial_\rho \Lambda^\mu_{\nu\sigma} - \partial_\sigma \Lambda^\mu_{\nu\rho} + \Lambda^\mu_{\alpha\rho} \Lambda^\alpha_{\nu\sigma} - \Lambda^\mu_{\alpha\sigma} \Lambda^\alpha_{\nu\rho} \quad (1.240)$$

We can obtain Ricci tensor by contracting with metric $g^{\mu\nu}$ as the following:

$$R_{\mu\nu} := R^\lambda_{\mu\lambda\nu} = g^{\lambda\sigma} R_{\sigma\mu\lambda\nu} \quad (1.241)$$

and Ricci scalar can be obtained by further contracting $R_{\mu\nu}$ as the following:

$$R = g^{\mu\nu} R_{\mu\nu} \quad (1.242)$$

To obtain the energy-momentum tensor $T^{\mu\nu}$, we can vary the matter action S_M under a metric transformation $g_{\mu\nu} \rightarrow g_{\mu\nu} + \delta g_{\mu\nu}$, in accordance to

$$\delta S_M = \frac{1}{2c} \int d^4x \sqrt{-g} T^{\mu\nu} \delta g_{\mu\nu} \quad (1.243)$$

The covariant derivative of a vector field $A^\mu(x)$ along the curve $x^\mu(\lambda)$ is defined as

$$\frac{DA^\mu}{D\tau} = \frac{dA^\mu}{d\tau} + \Gamma_{\nu\rho}^\mu A^\nu \frac{dx^\rho}{d\tau} \quad (1.244)$$

The Green's function is defined as

$$\square_x G(x - x') = \delta^4(x - x') \quad (1.245)$$

The various momenta of stress-tensor T^{ij} are

$$\begin{aligned} S^{ij}(t) &= \int d^3x T^{ij}(t, \mathbf{x}), \\ S^{ij,k}(t) &= \int d^3x T^{ij}(t, \mathbf{x}) x^k, \\ S^{ij,kl}(t) &= \int d^3x T^{ij}(t, \mathbf{x}) x^k x^l \end{aligned} \quad (1.246)$$

Useful Identities:

$$\begin{aligned} \int \frac{d\Omega}{4\pi} n_i n_j &= \frac{1}{3} \delta_{ij} \\ \int \frac{d\Omega}{4\pi} n_i n_j n_k n_l &= \frac{1}{15} (\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \end{aligned} \quad (1.247)$$

Chapter 2

Fast and faithful interpolation of numerical relativity surrogate waveforms using meshfree approximation

Abstract Several theoretical waveform models have been developed over the years to capture the gravitational wave emission from the dynamical evolution of compact binary systems of neutron stars and black holes. As ground-based detectors improve their sensitivity at low frequencies, the real-time computation of these waveforms can become computationally expensive, exacerbating the steep cost of rapidly reconstructing source parameters using Bayesian methods. This paper describes an efficient numerical algorithm for generating high-fidelity interpolated compact binary waveforms at an arbitrary point in the signal manifold by leveraging computational linear algebra techniques such as singular value decomposition and meshfree approximation. The results are presented for the time-domain `NRHybSur3dq8` inspiral-merger-ringdown (IMR) waveform model that is fine tuned to numerical relativity simulations and parameterized by the two component-masses and two aligned spins. For demonstration, we target a specific region of the intrinsic parameter space inspired by the previously inferred parameters of the `GW200311_115853` event – a binary black hole system whose merger was recorded by the network of advanced-LIGO and Virgo detectors during the third observation run. We show that the meshfree interpolated waveforms can be evaluated in ~ 2.3 ms, which is about $\times 38$ faster than its brute-force (frequency-domain tapered) implementation in the `PyCBC` software package at a median accuracy of $\sim \mathcal{O}(10^{-5})$. The algorithm is computationally efficient and scales favourably with an increasing number of dimensions of the parameter space. This technique may find use in rapid parameter estimation and source reconstruction studies.

2.1 Introduction

This chapter is based on the publication *Fast and faithful interpolation of numerical relativity surrogate waveforms using meshfree approximation*, arXiv:2403.19162 (2024).

Compact binary systems such as binary black holes or binary neutron stars are one of the most important sources for ground-based gravitational wave (GW) detectors. The GW signal emitted by such sources can be theoretically modeled in the secular inspiral or the post-merger ring-down phase by solving the Einstein's field equations. While analytical solutions are not available in the highly non-linear, so-called ‘merger’-domain of the evolution of these sources, the inspiral and ringdown solutions are often calibrated to a set of numerical relativity waveforms to construct complete semi-analytical waveforms covering the inspiral, merger and ringdown (IMR) phases of the evolution of such compact binary sources. Several such waveform models have been developed both in the frequency and time-domain, such as the family of frequency-domain IMR waveforms [77, 78, 79, 80, 81], and the EOB-family [82, 83] of waveforms, that are routinely used for GW data analysis. More recently, numerical relativity (NR) based surrogate models have also been developed such as the `NRSur7dq2` [84], `NRSur7dq4` [85], and the `NRHybSur3dq8` [86] models, which are among the most accurate waveform models available but are computationally expensive to generate in comparison to their frequency domain counterparts such as the `IMRPhenomXAS` and `SEOBNRv4_ROM` models. For reference, `NRHybSur3dq8` takes ~ 75 ms¹. to generate a waveform with component masses $m_{1,2} = (50, 20) M_\odot$ and aligned spin parameters $\chi_{1z,2z} = (0.05, 0.05)$ starting at a seismic cutoff frequency of 15 Hz.

Our research is inspired by a suggestion by Verma et al. [86], where the authors table the idea of speeding up the evaluation of the `NRHybSur3dq8` time-domain surrogate waveforms by creating a faster frequency-domain variant. In the past, fast frequency-domain surrogates using the technique of model order reduction have been applied to the `SEOBNRv4` [82] waveform model leading to a significant reduction in computational complexity of evaluating these waveforms. Such an approach could potentially accelerate parameter estimation (PE) for compact binary sources using these accurate waveform models.

There have been several attempts in the past to construct surrogate models of other waveform models in the past. Cannon et al. [87] constructed the approximate non-spinning IMR waveforms [88] parameterized by only the two-component masses - by first projecting these waveforms over a set of singular value decomposition (SVD) basis vectors, followed by a grid-based two-dimensional Chebyshev interpolation of the SVD coefficients. Chua et al. [89] used artificial neural networks (ANNs) to generate a four-dimensional reduced order model (ROM) of the 2.5 post-Newtonian (PN) frequency domain `TaylorF2` waveform model [90] parameterized by the two-component masses and the two aligned spins. In their approach, the waveforms are represented as the weighted sums over the reduced basis vectors. The ANNs are then trained to accurately map the GW source parameters into the basis coefficients. Other studies related to

¹The timing is for the implementation in the `PyCBC` software package

building NR surrogate models utilized ANNs [91], Gaussian Process Regression (GPR) [92], and deep learning [93] architectures.

In this work, we aim to significantly reduce the computational cost of generating frequency-domain `NRHybSur3dq8` waveforms by using a combination of the SVD-decomposition and meshfree approximation. We focus on illustrating our approach by concentrating on a region of the four-dimensional (two component masses and two aligned spins) parameter space around the inferred source parameters of the `GW200311_115853` [94] event detected during the third observing run of the LVK collaboration [95]. We construct meshfree interpolants for both the amplitude and phase of these waveforms separately. These interpolants are then used to rapidly evaluate the amplitude and phase of the waveform at any arbitrary query points in the chosen parameter space. These interpolated waveforms can be evaluated in ~ 2.3 ms in comparison to 89 ms ($\sim 38\times$ faster) with the standard (frequency-domain tapered) implementation of time-domain `NRHybSur3dq8` in `PyCBC` [15] with a median error of $\sim \mathcal{O}(10^{-5})$. Further, we also performed a Bayesian PE study of a simulated `GW200311_115853` like event coherently injected in Gaussian noise to mimic data from the network of LIGO and Virgo detectors. The PE runs were carried out using the meshfree interpolated waveforms and took ~ 16.4 minutes to complete on a 64 CPU-cores setup and the posterior distributions over the source parameters were found to be broadly consistent with their injected values.

The rest of the paper is organized as follows: Section 2.2 introduces the `NRHybSur3dq8` waveform model and the preprocessing required before we build meshfree interpolants. Section 2.3 explains an iterative strategy to find a suitable set of basis vectors using SVD to span the space of amplitude and phase and construct the meshfree interpolants of the resulting SVD coefficients. In Section 2.4, we describe various RBF kernels that can be used to generate meshfree interpolants. Section 2.5 demonstrates the results of a PE performed on a simulated BBH event using meshfree interpolants. Finally, we summarize the results in Section 2.6 and discuss the limitations of the current implementation and suggest some ideas for overcoming this limitation in follow-up studies.

2.2 NR Waveform model

`NRHybSur3dq8` is a time-domain, aligned spin surrogate model for “hybridized” non-precessing NR waveform, which is valid for stellar compact binaries with total masses as low as $2.5 M_{\odot}$. In this context, the term “hybridized” refers to a combined waveform incorporating both a post-Newtonian (PN) and effective one-body (EOB) waveform during the early times, attached smoothly to numerical relativity (NR) waveform at late times. The training of this model involves hybridized waveforms derived from 104 numerical relativity (NR) waveforms, supporting mass ratios $q \leq 8$, and spin values $\chi_{1z}; \chi_{2z} \leq 0.8$. Here, $\chi_{2z}(\chi_{2z})$ represents the spin of the heavier (lighter) black hole in alignment with the direction of orbital angular momentum. The waveform model also supports spin-weighted spherical harmonic modes with $l \leq 4$ and $(5, 5)$ mode but excludes the $(4, 0)$ or $(4, 1)$ modes. In this work, we specifically focus on the dominant, commonly referred to as the ”quadrupole” mode $(2, 2)$ present in this waveform model.

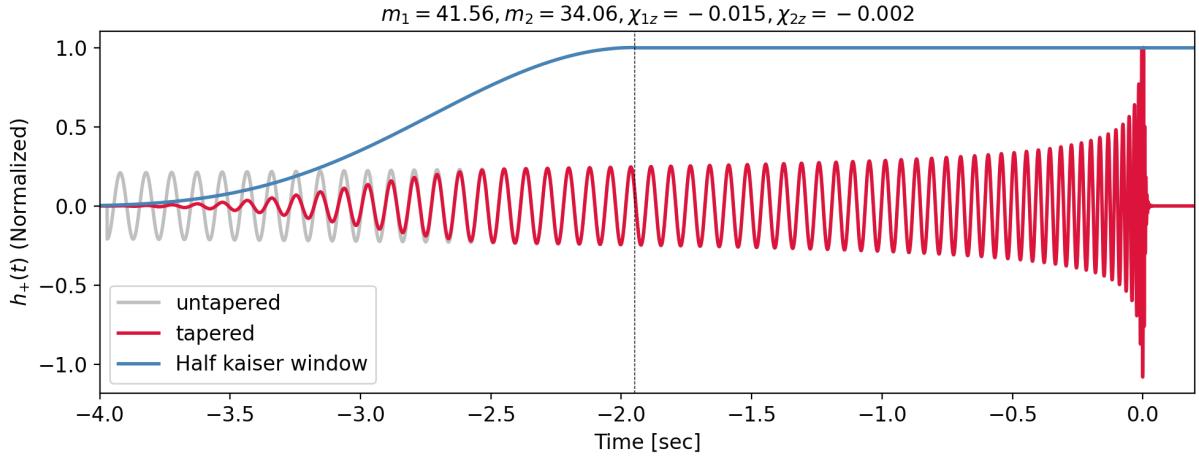


Figure 2.1: The NRHybSur3dq8 is generated at a frequency ($f_{start} = 10$ Hz) lower than the seismic cutoff frequency ($f_{low} = 15$ Hz), which leads to a longer duration waveform. It is generated for a simulated BBH event whose intrinsic parameters are shown in the title. A time-domain tapering using a half-Kaiser window is performed on the longer waveform to smoothly attenuate the waveform amplitude to zero, that helps in alleviating the Gibbs phenomenon arising due to a jump discontinuity while converting the time-domain waveform to a frequency-domain waveform. The black dashed vertical line represents the epoch at which the amplitude of the tapered and un-tapered waveform becomes equal (or equivalently, the value of the half-Kaiser window becomes 1). The sampling frequency is set at 2048 Hz.

The GW polarizations, denoted as “plus” (h_+) and “cross” (h_\times), can be concisely represented as a unified complex time series, denoted as $h = h_+ - i h_\times$. This complex time series can be expressed as a linear combination of spin-weighted harmonic modes h_{lm} [96, 97]. Consequently, the GW along any direction (ι, ϕ_0) in the binary source’s frame can be expressed as follows:

$$h(t, \iota, \phi_0) = \sum_{l=2}^{\infty} \sum_{m=-l}^l h_{lm}(t)^{-2} Y_{lm}(\iota, \phi_0) \quad (2.1)$$

Here, $^{-2}Y_{lm}$ denotes the spin $= -2$ weighted spherical harmonics, ι represents the inclination angle between the orbital angular momentum of the binary and the line of sight to the detector, and ϕ_0 corresponds to the initial binary phase. The corresponding frequency domain waveform for the dominant mode can be written as the following:

$$\tilde{h}_{+/\times} = \mathcal{A}_{+/\times}(f) \exp[j \psi_{+/\times}(f)] \quad (2.2)$$

where $\mathcal{A}_{+/\times}(f)$ and $\psi_{+/\times}(f)$ are amplitude and phase as a function of frequency, respectively. Since NRHybSur3dq8 is an aligned-spin waveform model [98, 99, 100, 101, 102, 103], the relation $\tilde{h}_\times \propto -i \tilde{h}_+$ holds and therefore we only consider \tilde{h}_+ and the cross-polarization ($\tilde{h}_\times(f)$) can be calculated from the above relation. From now onwards, we will drop the subscript ‘+/\times’ from $\mathcal{A}_{+/\times}$ and $\psi_{+/\times}$ and simply denote the amplitude and phase by \mathcal{A} and ψ which correspond to plus polarization.

As a pre-processing step, we first calculate the duration of the waveform starting at the seismic cutoff frequency (f_{low}). Subsequently, the starting frequency (f_{start}) of the waveform

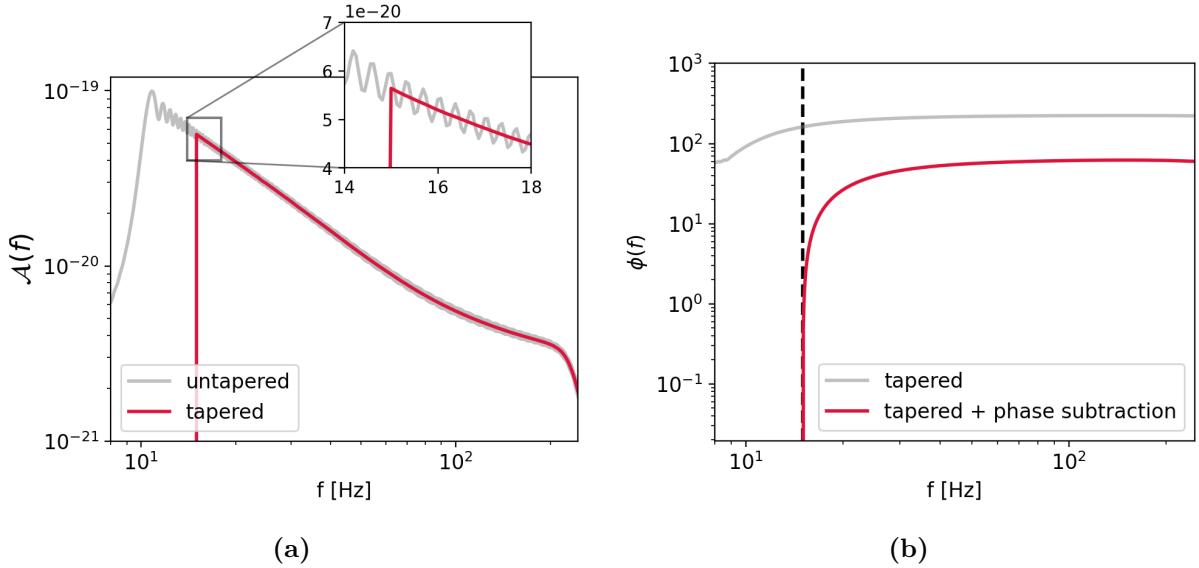


Figure 2.2: *left:* The amplitude of the untapered waveform (*solid red*) is plotted with that of the tapered (*solid grey*) waveform. Note the ringing artifacts appearing due to the Gibbs phenomenon in the untapered waveform. It is almost negligible in the amplitude of the tapered waveform. *right:* After converting the tapered waveform into the frequency domain, there is some accumulated phase below $f_{low} = 15$ Hz, which we subtract from the total phase so that the phase of the frequency domain waveform is zero at $f_{low} = 15$ Hz. The black-dashed vertical line represents $f_{low} = 15$ Hz.

generation is decreased until we reach twice the duration of the original waveform. Then the time-domain NRHybSur3dq8 waveform is generated starting at f_{start} followed by a time-domain tapering to smoothly decrease the amplitude to zero (see Fig. 2.1). The length of the tapering window is taken as a fraction (0.8 in this case) of the difference between the new and original duration of the waveform. The longer duration of the waveform makes sure that we do not lose any portion of the waveform in the frequency band to be used in the PE while tapering. We use a Kaiser window [104] (as implemented in PyCBC) for time-domain tapering, which can be expressed as

$$w(n) = I_0 \left(\beta \sqrt{1 - \frac{4n^2}{(M-1)^2}} \right) / I_0(\beta) \quad (2.3)$$

with $n \leq \lfloor \frac{M-1}{2} \rfloor$ and where I_0 is the modified zeroth-order Bessel function, M is the number of points in the output window, and β is the shape parameter, which determines the trade-off between main-lobe width and side lobe level in the Fourier response of the window. As the beta becomes large, the main lobe width increases while the side lobe level decreases. For this analysis, we use a half-Kaiser window for tapering (see Fig. 2.1). The tapering helps in reducing the Gibbs phenomenon while converting the tapered waveform in the frequency domain (see Fig. 2.2a). Additionally, we also subtract the phase accumulated up to f_{low} from the total phase of the frequency domain waveform (tapered) to make sure the phase at f_{low} is zero, which significantly helps in the interpolation of the SVD coefficients corresponding to the phase (see Fig. 2.2b).

In the next section, we lay down an iterative strategy to find suitable basis vectors to span the

space of amplitude and phase using SVD. Subsequently, we fit the resulting SVD coefficients for amplitude and phase separately using a linear combination of RBFs and monomials. Since the amplitude and phase are smoothly varying functions of frequency (see Fig. 2.2a and Fig. 2.2b), the corresponding SVD coefficients are expected to exhibit smooth variation over the intrinsic parameter space as well and, therefore, suitable for interpolation. Combining the interpolated coefficients at the arbitrary query points within the interpolating region with the corresponding basis vectors gives the interpolated amplitude and phase and, hence, the interpolated waveform.

2.3 Generation of meshfree interpolants

To construct the interpolants for evaluating the meshfree waveforms, the initial step involves choosing a patch of intrinsic parameter space for which interpolants are to be built. The dimensionality of the patch depends on the dimensionality of the intrinsic parameter space over which interpolation is to be performed. Rather than opting for a single patch with large boundaries, a more effective approach (as suggested in [105]) is to create multiple overlapping patches with smaller boundaries to comprehensively encompass the desired parameter space. For each patch, meshfree interpolants are independently constructed, enhancing the accuracy of individual interpolants. In this analysis, our selection of patches is inspired by a simulated BBH event with `GW200311_115853`-like parameters. We use `NRHybSur3dq8`, an aligned spin numerical relativity (NR) surrogate waveform model, which also includes support for subdominant modes as well. As a proof of principle study, this analysis focuses only on the leading-order mode; however, the procedure presented here can be extended to incorporate higher-order modes. The procedure to generate interpolants comprises the following steps:

2.3.1 Selection of patch(es)

This task involves identifying the patch (or patches) within the parameter space for the construction of waveform interpolants². While it could be any segment of the pertinent parameter space in general, for this study, we opt for a patch that encompasses the injected values of parameters associated with the simulated BBH event. We define two four-dimensional hyper-rectangular patches with the following ranges: $\mathcal{M} \in [25, 40]$, $q \in [1, 3]$, $\chi_{1z} \in [-0.1, 0.1]$, and $\chi_{2z} \in [-0.05, 0.05]$. Both patches share common ranges for \mathcal{M} , q , and χ_{2z} . However, for the first patch, $\chi_{1z} \in [-0.1, 0]$, and for the second patch, $\chi_{1z} \in [-0.005, 0.1]$. Note that these two patches overlap in the χ_{1z} dimension (see Fig. 2.3a).

²A similar approach was taken by Morisaki et al. in their work [106] for constructing ROM bases corresponding to different patches of parameter space.

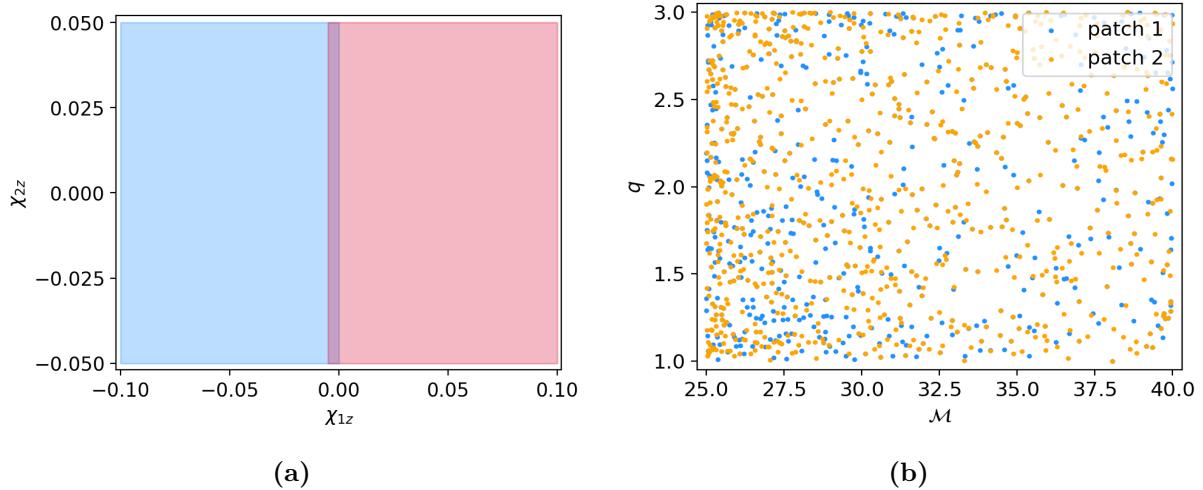


Figure 2.3: *left:* We choose two overlapping patches over χ_{1z} to cover the desired range in χ_{1z} . A similar procedure can be performed to extend the range of the other parameters (say \mathcal{M} , q etc.). *right:* Final nodes after iterative SVD process terminates for both patches. The nodes found using the “Iterative SVD” procedure tend to cluster near the lower mass boundaries.

2.3.2 Placing RBF nodes using ‘Iterative SVD’ method

This step starts by randomly spraying a minimum³ number of nodes (N) across the four-dimensional sample space and generating frequency domain waveforms (as specified in Sec. 2.2) at these specified nodes. Subsequently, \mathcal{A} and ψ are extracted from $\tilde{h}_+(f)$ for each of the N waveforms. The obtained amplitudes (\mathcal{A}) and phases (ψ) are then stacked in a row-wise fashion to construct two matrices, \mathbf{A} and Ψ . These matrices are subjected to singular value decomposition (SVD), which, in turn, provides the relevant basis for spanning the space defined by these amplitudes and phases.

$$\mathcal{A} = \sum_{\mu=1}^N C_{\mu}^{\mathcal{A}} \vec{u}_{\mu}^{\mathcal{A}} \quad \text{and} \quad \psi = \sum_{\mu=1}^N C_{\mu}^{\psi} \vec{u}_{\mu}^{\psi} \quad (2.4)$$

Since by construction, the basis vectors $(\vec{u}_{\mu}^{\mathcal{A}}, \vec{u}_{\mu}^{\psi})$ are in the decreasing order of importance, it turns out that we need to retain only top few basis vectors (say upto ℓ) to accurately reconstruct amplitude and phase as illustrated by the singular values spectrum in Fig. 2.4. Consequently, we limit the construction of interpolants to the SVD coefficients $C_{\mu}^{\mathcal{A}}$ and C_{μ}^{ψ} corresponding to top- ℓ basis vectors. Later on, we can choose to retain an even lesser number of basis vectors while reconstructing amplitude and phase at arbitrary query points. These coefficients, denoted as $C_{\mu}^{\mathcal{A}}$ and C_{μ}^{ψ} , exhibit smooth variations as functions of $\vec{\lambda}$, where $\vec{\lambda} \equiv \{\mathcal{M}, q, \chi_{1z}, \chi_{2z}\}$ represents the interpolating parameter space. These coefficients can be written as a linear combination of

³The initial number of nodes to start the algorithm can be decided by the degree of the monomial (see Eq. (2.5b)) and dimensionality of the intrinsic parameter space.

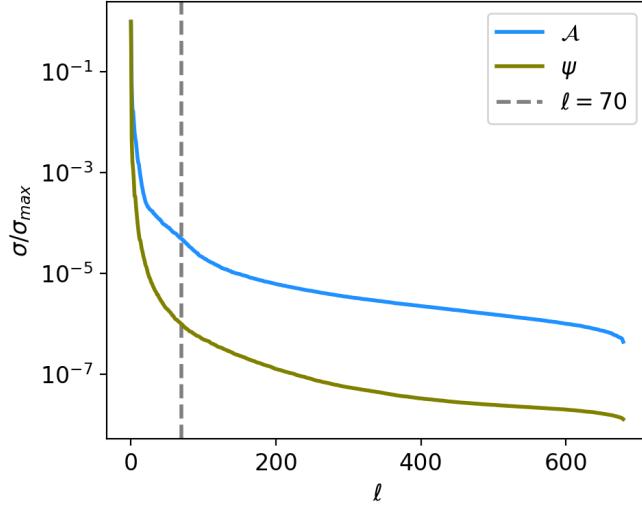


Figure 2.4: Singular values spectrum corresponding to amplitude \mathcal{A} and phase φ . Note that singular values drop to $\sim \mathcal{O}(10^{-6})$ within ~ 70 basis vectors.

RBFs (ϕ) and monomials (p) of specified degree as follows [107]:

$$C_{\mu}^{\mathcal{A}(q)} = \sum_{n=1}^N a_n^{\mathcal{A}} \phi(\|\vec{\lambda}^q - \vec{\lambda}^n\|_2) + \sum_{j=1}^M b_j^{\mathcal{A}} p_j(\vec{\lambda}^q) \quad (2.5a)$$

$$C_{\mu}^{\psi(q)} = \sum_{n=1}^N a_n^{\psi} \phi(\|\vec{\lambda}^q - \vec{\lambda}^n\|_2) + \sum_{j=1}^M b_j^{\psi} p_j(\vec{\lambda}^q) \quad (2.5b)$$

The addition of these monomial terms enhances the accuracy of RBF interpolation, as shown in [108]. Since these coefficients are known at the N nodes, we can invert the above equations to calculate the coefficients $a^{\mathcal{A}}$, a^{ψ} , $b^{\mathcal{A}}$, and b^{ψ} corresponding to each basis vector. Using these coefficients, we can quickly evaluate the $C_{\mu}^{\mathcal{A}(q)}$ and $C_{\mu}^{\psi(q)}$ (for μ^{th} basis) at an arbitrary query point within the interpolating parameter space and then combine it with corresponding basis vectors ($\vec{u}^{\mathcal{A}}$ and \vec{u}^{ψ}) to get back the interpolated amplitude \mathcal{A} and phase ψ . Then, we combine both amplitude and phase in accordance with Eq. (2.2) to get the frequency domain waveform.

Once we have the interpolants ready, a set of 10^3 query points is generated within the interpolating patch, and corresponding true frequency domain waveforms and meshfree interpolated waveforms are generated at these query points. Subsequently, the mismatches between the true and interpolated waveforms are calculated. A specified number of query points (10 for this work) exhibiting the worst mismatch are added back into the original set of initial nodes. The process then reverts to Subsection 2.3.2 and continues until either a limit of maximum nodes is reached or the maximum mismatch at the current iteration falls below a predefined threshold (e.g., $\leq 10^{-4}$) on maximum mismatch ⁴. We define the match between the

⁴Note that the mismatches are calculated between meshfree waveforms and waveforms first generated in time-domain at $f_{start} = 10$ Hz, then tapered and converted into frequency domain followed by phase subtraction as in Section 2.2.

two waveforms h_1 and h_2 as

$$\mathcal{O}(h_1, h_2) = \frac{\langle h_1 | h_2 \rangle}{\sqrt{\langle h_1 | h_1 \rangle \langle h_2 | h_2 \rangle}} \quad (2.6)$$

where,

$$\langle h_1 | h_2 \rangle = \int_{f_{\text{low}}}^{f_{\text{high}}} \frac{\tilde{h}_1(f)\tilde{h}_2^*(f)}{S_n(f)} df \quad (2.7)$$

which defines a noise-weighted inner product with PSD, S_n and $*$ denotes the complex conjugation. The match is also maximized over a time and phase shift between two waveforms. Finally, the mismatch is defined as $1 - \mathcal{O}(h_1, h_2)$.

We call this process “Iterative SVD” because it progressively refines the meshfree waveform approximation by reinserting the query points with the highest mismatch back into the initial set of nodes, followed by evaluation of the true frequency domain waveforms at the appended set of nodes. Currently, with each iteration of the “iterative SVD” node placement algorithm, as additional RBF nodes are introduced, the data matrices \mathbf{A} and Ψ are expanded by incorporating more rows (one for each newly added node) followed by a Singular Value Decomposition (SVD) conducted on these expanded data matrices. This approach is practical for handling the decomposition of the small-sized data matrices. However, one can update the basis vectors and singular values at each iteration with more sophisticated algorithms [109, 110] that update the left and right subspace and singular values from the existing SVD decomposition of a dense matrix as new rows are added to it. For updating the subspaces and singular values, Brand et al. [109] developed an algorithm for computing a thin SVD of a data matrix of size $(m \times n)$ in a single pass with linear time complexity $\mathcal{O}(mnr)$, where $r \leq \sqrt{\min(m, n)}$. The authors proposed fast and memory-efficient sequential algorithms for tracking the singular values and subspaces, initialized with a general identity included in the existing decomposing of the data matrix. Adding a new column would modify the identity representation and provide rank one updates via the modified Gram–Schmidt algorithm.

Note that if the required number of RBF nodes is small, changing the strategy for computation of the subspace and coefficients using the “updated SVD” approach would not significantly reduce the overall computational cost. On the other hand, for a large number of nodes, the incorporation of such sophisticated linear algebra algorithms in our proposed framework would be beneficial. We will explore this possibility in detail in future work.

Next, the SVD is performed for the new amplitude and phase matrices. This is a similar approach that was used in previous studies for building reduced-order models (ROMs) of a number of waveform approximants [111, 112, 113, 106, 114]. In each iteration, SVD is employed to identify improved basis vectors for \mathcal{A} and ψ . As shown in Fig. 2.5, the maximum mismatch calculated between the worst meshfree waveform and the corresponding true waveform decreases as we add the query points corresponding to the worst meshfree waveforms at each iteration into the set of nodes used to generate the interpolants. The calculation of mismatches assumes a `aLIGOZeroDetHighPower` [115] Power Spectral Density (PSD) as implemented in PyCBC. We repeat the same “Iterative SVD” procedure for the other interpolating patches (if needed). For the analysis in this paper, we choose two patches (see Section 2.5). As evident from Fig. 2.3b,

RBF nodes (for both patches) found by the ‘‘Iterative SVD’’ tend to cluster more towards the low-mass boundaries implying that the points near the lower mass boundaries contribute the most in terms of the suitable basis vectors for each patch. Similar patterns are also seen in the work focusing on building a surrogate of PN waveforms [86] to find the desired training dataset of the parameters and building ROM using ANNs [116], using a greedy strategy. Similar patterns

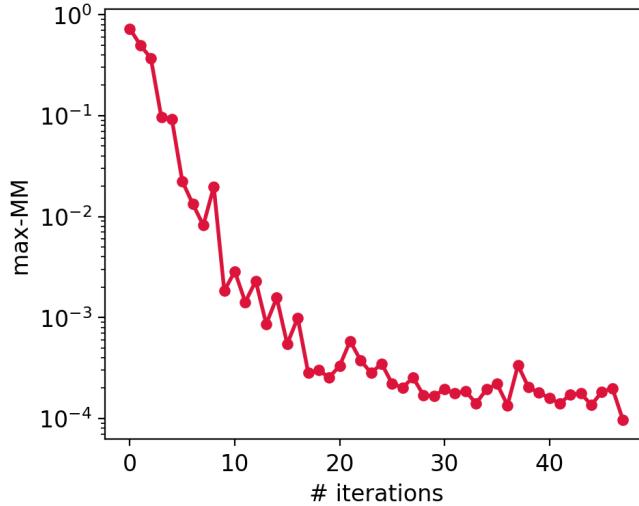


Figure 2.5: Mismatch between the worst approximated meshfree waveform and true waveform at each iteration. As evident from the figure, with each iteration, the meshfree approximation becomes better.

are also seen in the work focusing on building a surrogate of PN waveforms [86] to find the desired training dataset of the parameters and building ROM using ANNs [116], using a greedy strategy.

Note that initially, the maximum mismatch at each iteration is decreasing rapidly. After ~ 25 iterations, the rate of decrease in the mismatch reduces despite adding 10 worst points at each iteration. It implies that it is close to having the minimum number of nodes required to satisfy the maximum mismatch threshold criteria mentioned earlier. After ~ 48 iterations, the maximum mismatch falls below 10^{-4} , and algorithm stops. However, to confirm whether it has indeed crossed this criterion (since there are fluctuations), we can test the accuracy of interpolants again by spraying a different set of random query points within the interpolating parameter space and confirm that the maximum mismatch is still less than 10^{-4} . Otherwise, we can again start the algorithm until this threshold criterion is met by at least two sets of random query points.

2.4 Choice of RBF kernel

There are different RBF kernels that can be used as a basis (see Eq. (2.5b)) for meshfree approximation. In this work, we tried the following three different RBF kernels (see table 2.1): (i) Inverse multiquadric ('imq'), (ii) Inverse Quadric ('iq'), and (iii) Gaussian ('ga'). As mentioned earlier in sec. 2.3, to terminate the iterative SVD, we set a threshold of either a maximum

number of nodes (3000 here) or a maximum mismatch of 10^{-4} or less at a given iteration. The

RBF kernel type	$\phi(r)$	ϵ	
		Patch 1	Patch 2
inverse multiquadric ('imq')	$1/\sqrt{1 + (\epsilon r)^2}$	0.4016	0.4010
Gaussian ('ga')	$\exp[-(\epsilon r)^2]$	0.3172	0.2909
inverse quadric ('iq')	$1/(1 + (\epsilon r)^2)$	0.4464	0.4378

Table 2.1: Different RBF kernels

interpolant generation is stopped whenever any of the above-stated conditions is satisfied. In the case of 'ga'('iq') kernels, the final number of nodes for patch 1 and patch 2 was 1530(900) and 1620(590), respectively, when the maximum mismatch at an iteration fell below the threshold of 10^{-4} . However, in the case of 'imq' RBF kernel, the final number of nodes for patch one was 680, while for the second patch, it was 770 for the same threshold. A natural question arises: what is so special about the 'imq'? A possible explanation is the spread of these RBF kernels. As shown in Fig. 2.6, the 'ga', 'imq', and 'iq' have similar profiles near the center ($r = 0$). However, as we go away from the center, the 'ga' and 'iq' fall very rapidly in comparison to the 'imq' kernel. It implies that the 'imq' kernel has a significant overlap with the 'imq' kernel centered at the other nearby nodes (centers) in comparison to the other two RBF kernels, and therefore, it explains the lesser number of nodes required to fall below the maximum mismatch threshold at a given iteration for 'imq' in comparison to a relatively higher number of nodes for the other RBF kernels. For the next section, we choose 'imq' as the RBF kernel for generating meshfree interpolants.

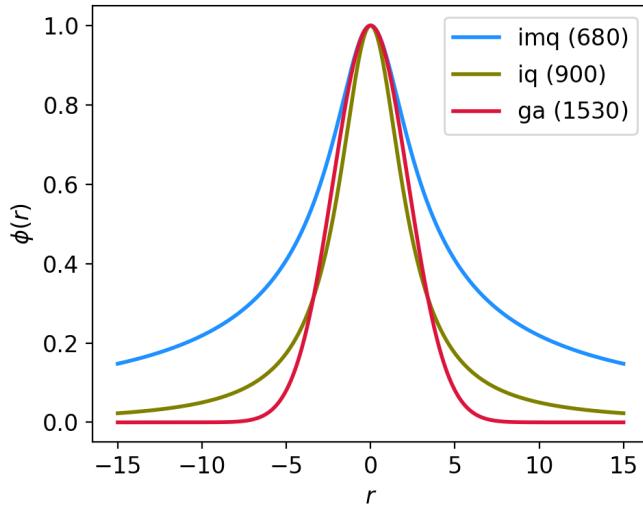


Figure 2.6: Profiles of different RBF kernels. In the legend, numbers in the bracket are the final number of nodes required (for patch 1) to fall below the maximum mismatch threshold of 10^{-4} in the iterative SVD process. For the 'imq' kernel, we required the least number of nodes to achieve the threshold.

2.5 Results

As previously mentioned, there are two overlapping patches in χ_{1z} , and we independently construct interpolants for each of the patches. In the first patch where $\chi_{1z} \in [-0.1, 0]$, we initiate with $N = \binom{\nu+d}{\nu}$ nodes, with ν representing the degree of the monomial terms in Eq. (2.5b) and d denoting the dimensionality of the parameter space (here, $\nu = 6$ and $d = 4$ for the four-dimensional parameter space interpolation). The ‘imq’ RBF kernel is selected, where ϵ is the shape factor influencing the RBFs’ spread at the nodes, determined through (LOOCV) procedure [117] (see table 2.1 for values of ϵ). In the second patch covering $\chi_{1z} \in [-0.005, 0.1]$, the RBF parameters chosen for the first patch are retained, with the exception of a different ϵ value for this patch. The seismic cutoff frequency (f_{low}) is set at 15 Hz while f_{start} is equal to 10 Hz. Once the interpolants for both patches are ready (see Sec. 2.3), we can use them to generate the interpolated waveforms at the arbitrary query points within the interpolating sample space. Note that the generation of interpolants is a one-time offline process, and it can be completed well in advance before the parameter estimation of the GW event is initiated. While multiple patches are necessary to adequately cover the desired parameter space ranges, the generation of interpolants is highly parallelizable and can benefit from a multicore setup to expedite this stage. In the specific case of this analysis, the generation of interpolants for each patch was

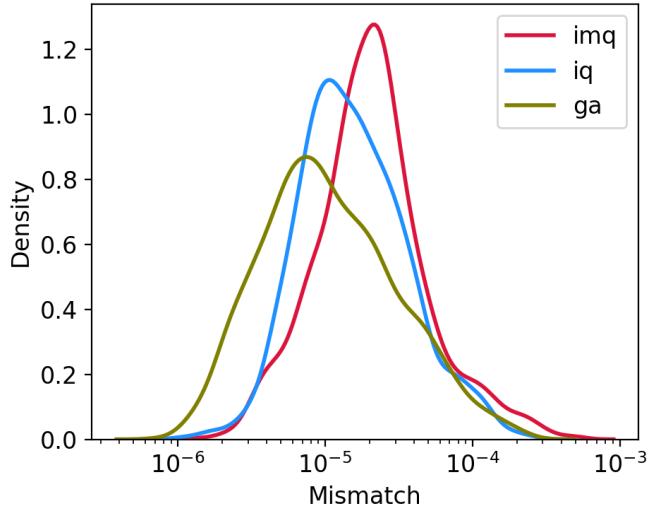


Figure 2.7: The mismatches calculated between interpolated and true waveforms at the posterior samples are shown. For comparison, we also show the mismatches for meshfree waveforms using ‘ga’ and ‘iq’ RBF kernels. The median mismatched (for all kernels) is $\mathcal{O}(10^{-5})$. Note that the accuracy achieved with ‘ga’ is a little better than with the other kernels due to the high number of nodes that were used to construct ‘ga’ kernel-based meshfree interpolants. However, we do not find any significant difference in the corresponding posterior distributions.

accomplished in ~ 20 minutes using a CPU setup with 32 cores each. We use the publically available RBF Python package [118] to generate RBF interpolants.

To assess the accuracy of the interpolated waveforms, we generated 20 sets of 10^3 query points

each, randomly distributed in the intrinsic parameter space within the interpolating region (see table 2.2). We then generated both the true and interpolated waveforms at these query points and computed the mismatches between them, assuming a flat Power Spectral Density (PSD). The median mismatch is found to be $\mathcal{O}(10^{-5})$, which shows the high level of accuracy of the meshfree waveforms in approximating the true waveforms across the interpolating sample space

Parameters	Range	Prior distribution
\mathcal{M}	[25, 40]	$\propto \mathcal{M}$
q	[1, 3]	$\propto [(1+q)/q^3]^{2/5}$
χ_{1z}	[-0.1, 0.1]	Uniform
χ_{2z}	[-0.05, 0.05]	Uniform
V_{com}	[5e3, 1e11]	Uniform
t_c	$t_{\text{trig}} \pm 0.12$	Uniform
α	[0, 2π]	Uniform
δ	$\pm\pi/2$	$\sin^{-1}[\text{Uniform}[-1, 1]]$
ι	[0, π]	Uniform in $\cos \iota$
ψ	[0, 2π]	Uniform angle

Table 2.2: Prior parameter space over the ten-dimensional parameter space $\vec{\Lambda}$.

In terms of the computational speed-ups, the meshfree waveforms can be evaluated in ~ 2.3 ms in comparison to ~ 89 ms⁵ implying a speed-up factor of ~ 38 . Note that these timings are for the Python implementation of these waveforms. A “C implementation” of the true waveforms can be evaluated in ~ 10 ms as shown by [86] (about $4.3\times$ slower than the meshfree waveform’s Python implementation). Since these numbers are also dependent on the hardware configurations of the machines on which these speed tests are performed, a fair comparison of meshfree speed-ups with their C implementation is not possible until we make a C version of meshfree waveforms.

Finally, to test the accuracy of the meshfree model in the context of parameter estimation, we use `SEOBNRv4_ROM` [82] waveform model to inject 16 seconds long simulated BBH event into the Gaussian noise with PSDs taken from `aLIGOZeroDetHighPower` [115] for both LIGO-Livingston and LIGO-Hanford detectors, and `AdvVirgo` [31] for Virgo detector with network matched-filter signal-to-noise ratio (SNR) ~ 18 . The seismic cutoff frequency (f_{low}) is set at 15 Hz, and the high cutoff frequency is equal to the ringdown frequency of the lowest component masses within the chosen range of chirp masses and mass ratios. A sampling frequency of 2048 Hz is considered. We employ `dynesty` [72, 74], sampler, a Python implementation of the Nested sampling algorithm, to sample the posterior distribution. The prior distributions and the boundaries of the parameters (to be estimated) are shown in Table 2.2. We choose the following sampler settings for `dynesty`: `nLive` = 500, `nWalks` = 500, `dlogz` = 0.1, `sample` = “rwalk”. Here,

⁵frequency domain tapered version of `NRHybSur3dq8` time-domain model

the parameter `nlive` represents the number of live points, which determines the resolution of the sampled posterior distribution.

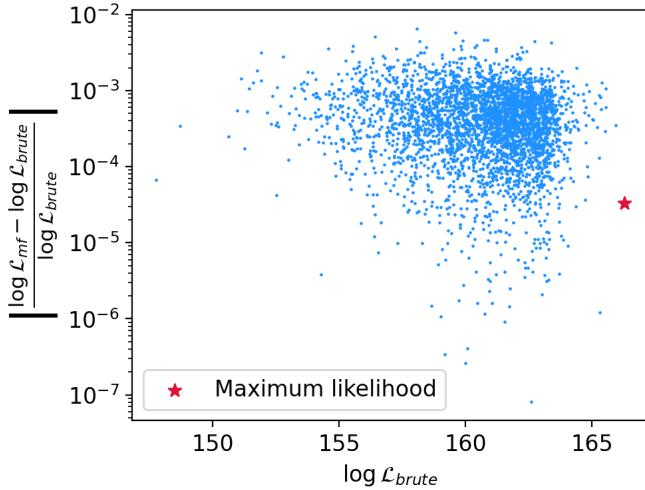


Figure 2.8: Absolute relative errors in the likelihood as a function of true $\log \mathcal{L}$. The absolute relative error is defined as $|(\log \mathcal{L}_{\text{mf}} - \log \mathcal{L}_{\text{brute}}) / \log \mathcal{L}_{\text{brute}}|$. The $\log \mathcal{L}_{\text{mf}}$ is calculated using the interpolated waveforms, while $\log \mathcal{L}_{\text{brute}}$ is calculated using the true waveforms. These likelihoods were calculated at the posterior samples obtained from the PE of the simulated BBH event. The red star represents the maximum likelihood posterior sample. Note that the majority of relative errors (especially near the peak) are well within 1%, demonstrating the accuracy of the interpolants.

Smaller values of `nLive` might give rise to a poorly sampled distribution (hence evidence) with much faster convergence. Instead, taking a larger value can give us a finely sampled distribution at the expense of lower convergence speeds. `walks` specifies the minimum number of steps required before a new live point is proposed, which replaces the live point with the lowest likelihood in the nested sampling. `sample` indicates the chosen approach for generating samples, and `dlogz` characterizes the remaining prior volume's contribution to the total evidence. In this PE study, the sampling terminates once `dlogz` reaches a threshold of 0.1. For a more comprehensive understanding of `dynesty`'s nested sampling algorithm and its practical implementation, one can refer to the following references [72, 73]. As evident from Fig. 2.9, posterior distributions of various binary parameters of simulated BBH event contain the injected values well within the 90% CI. This PE analysis took ~ 16.4 minutes on a 64 CPU cores setup. Fig. 2.7 also shows the probability distribution function (PDF) of the mismatches of the interpolated waveforms with the true waveforms generated at the posterior samples obtained from the PE and median mismatch is $\sim \mathcal{O}(10^{-5})$, demonstrating the good accuracy of the meshfree waveforms. To quantify the effect of these approximate waveforms on the likelihood calculation, we also evaluated the relative errors in the likelihood, as shown in Fig. 2.8 and found a median absolute relative error of $\sim \mathcal{O}(10^{-3})$ further showing the effectiveness of the meshfree waveforms. All the tests were performed on AMD EPYC 7542 CPU@2.90GHz processors.

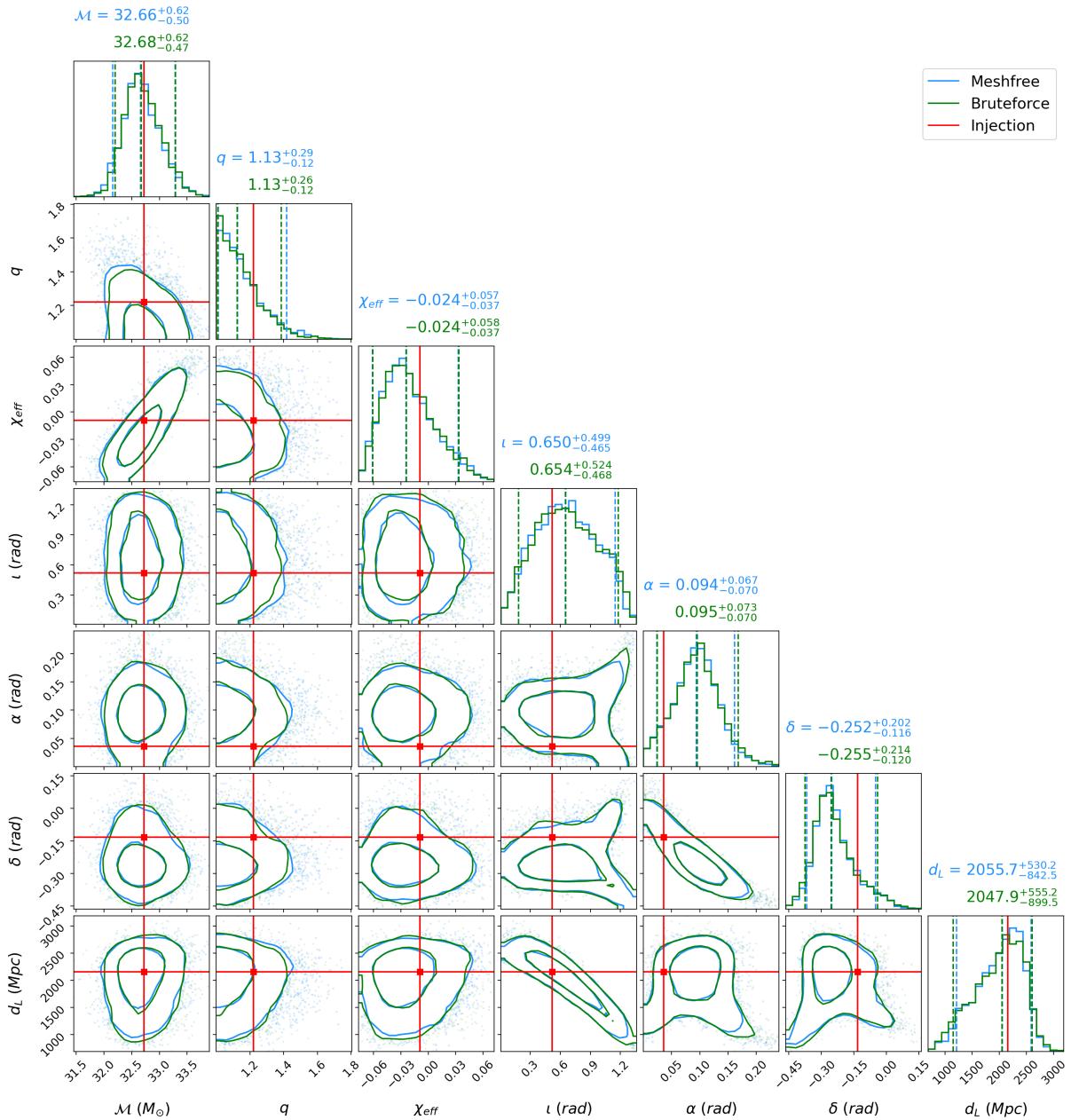


Figure 2.9: Corner plot of the posterior distributions (with 50% and 90% CI and contours) obtained by the meshfree interpolated waveforms for a simulated BBH event. The vertical red line represents the injected values. It took ~ 16.4 min to complete on 64 CPU cores. In comparison, the PE run with the direct evaluation of the frequency-domain version of the tapered time-domain NRHybSur3dq8 waveform model took ~ 5 h and 50 min on the same number of CPU cores and the resulting posterior distributions are broadly consistent with those obtained using meshfree interpolated waveforms.

2.6 Conclusion and Future Outlook

In this work, we have constructed a fast, interpolated frequency domain model that approximates the time-domain NRHybSur3dq8 waveform model with high fidelity. We show that such interpolated waveforms can be evaluated in ~ 2.3 ms ($\times 38$) times faster than the standard

implementation of this waveform model in PyCBC. We constructed the waveform interpolants by first selecting a patch of the parameter space. Then, we spray randomly scattered nodes in the sample space and evaluate \mathcal{A} and ψ from the frequency domain waveform corresponding to `NRHybSur3dq8` at those nodes. Subsequently, we find a suitable basis spanning the space of both amplitude and phase by performing the SVD of the amplitude and phase matrices resulting from stacking the amplitudes and phases at the nodes. The resulting SVD coefficients corresponding to amplitude and phase are fit with a linear combination of RBFs and monomials (“interpolants”), followed by the mismatched calculation of the interpolated waveforms with the true waveforms at arbitrary query points in the selected patch of parameter space. The worst 10 query points in terms of mismatch are added back into the set of initial nodes, and the whole process (called “Iterative SVD”) repeats till either the maximum mismatch at the given iteration falls below a set threshold, or the number of nodes crosses a fixed number of maximum nodes. In each iteration, the accuracy of the interpolants increases, which in turn enhances the meshfree approximation of the waveforms. Note that the choice of 10 worst points, as mentioned earlier, is completely arbitrary and chosen to accelerate the interpolants generation process. In order to test the validity of the meshfree interpolated waveforms, we also performed a parameter estimation of a simulated BBH event using the meshfree interpolated waveforms and found posterior distributions are consistent with the injected values. It is important to highlight that our method does not require importance sampling as used in deep learning methods such as DINGO [119] and NESSAI [120].

Currently, a limitation we face involves the multiplication of SVD coefficients with the basis vectors. As the duration of the waveform increases, the computational expense of this multiplication rises, consequently impacting the evaluation cost of the meshfree interpolated waveform. In this context, we can also use the Empirical interpolation method (EIM) [121, 122, 123, 112] to reduce the number of frequency points included in the basis vectors. It is yet to see whether this kind of strategy could be helpful within the meshfree framework in reducing the dimensionality of the basis vectors, which is especially needed for longer-duration waveforms. Further, in this work, we used an iterative SVD scheme that updates the basis vectors with a new set of points by applying independent SVD of the updated amplitude and phase matrix. In the follow-up work, we will replace our iterative SVD scheme with advanced updated SVD algorithms [109, 110] and will compare the computation cost of performing independent SVD in each iteration and updating the initial basis and singular values from the first iteration using updated SVD algorithms. Another notable aspect is that the distance employed in Eq.(2.5b) is Euclidean, even though the parameter space we operate in is inherently non-Euclidean. Nevertheless, we choose nodes in chirp-time coordinates (θ_0 and θ_3) instead of component masses or chirp mass and mass ratio as the metric varies slowly in chirp-time coordinates. Additionally, it’s worth noting that this work solely focuses on the aligned spin waveform model `NRHybSur3dq8`. Subsequent extensions of this work will explore broadening the meshfree framework to incorporate NR surrogate models encompassing subdominant modes, considering spins, and addressing the aforementioned limitations. We also performed similar exercises for other aligned spin waveform models, such as `IMRPhenomXAS` and `SEOBNRv4_ROM`.

Since these waveform models are already fast to evaluate, we don't get any significant speed-up in their corresponding meshfree waveform evaluation. However, the speed-up can be further enhanced by employing strategies such as adaptive frequency resolution [124] and EIM procedure, as mentioned earlier. In future follow-ups of this work, we would also consider extending this meshfree framework to include NR surrogate models that include subdominant modes, precessing spins, and possibly eccentric models.

Acknowledgements

We thank Srashti Goyal for carefully going through the manuscript and giving helpful comments. We thank Prayush Kumar for his help in the initial stages of this work. We also thank Abhishek Sharma and Sachin Shukla for their useful suggestions and comments. L. P. is supported by the Research Scholarship Program of Tata Consultancy Services (TCS). A. R is supported by the research program of the Netherlands Organisation for Scientific Research (NWO). A. S. gratefully acknowledges the generous grant provided by the Department of Science and Technology, India, through the DST-ICPS (Interdisciplinary Cyber Physical Systems) cluster project funding. We thank the HPC support staff at IIT Gandhinagar for their help and cooperation. The authors are grateful for the computational resources provided by the LIGO Laboratory and supported by the National Science Foundation Grants No. PHY-0757058 and No. PHY-0823459. This material is based upon work supported by NSF's LIGO Laboratory, which is a major facility fully funded by the National Science Foundation.

This research has made use of data or software obtained from the Gravitational Wave Open Science Center [125], a service of the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA. This material is based upon work supported by NSF's LIGO Laboratory, which is a major facility fully funded by the National Science Foundation, as well as the Science and Technology Facilities Council (STFC) of the United Kingdom, the Max-Planck-Society (MPS), and the State of Niedersachsen/Germany for support of the construction of Advanced LIGO and construction and operation of the GEO600 detector. Additional support for Advanced LIGO was provided by the Australian Research Council. Virgo is funded through the European Gravitational Observatory (EGO), the French Centre National de Recherche Scientifique (CNRS), the Italian Istituto Nazionale di Fisica Nucleare (INFN), and the Dutch Nikhef, with contributions by institutions from Belgium, Germany, Greece, Hungary, Ireland, Japan, Monaco, Poland, Portugal, Spain. KAGRA is supported by the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan Society for the Promotion of Science (JSPS) in Japan; National Research Foundation (NRF) and Ministry of Science and ICT (MSIT) in Korea; Academia Sinica (AS) and National Science and Technology Council (NSTC) in Taiwan.

Code availability: Codes used in this analysis are publicly available in a Github repository [126].

Chapter 3

Fast likelihood evaluation using meshfree approximations for reconstructing compact binary sources

Abstract Several rapid parameter estimation methods have recently been advanced to deal with the computational challenges of the problem of Bayesian inference of the properties of compact binary sources detected in the upcoming science runs of the terrestrial network of gravitational wave detectors. Some of these methods are well-optimized to reconstruct gravitational wave signals in nearly real-time, necessary for multi-messenger astronomy. In this context, this work presents a new, computationally efficient algorithm for fast evaluation of the likelihood function using a combination of numerical linear algebra and mesh-free interpolation methods. The proposed method can rapidly evaluate the likelihood function at any arbitrary point of the sample space at a negligible loss of accuracy and is an alternative to the grid-based parameter estimation schemes. We obtain posterior samples over model parameters for a canonical binary neutron star system by interfacing our fast likelihood evaluation method with the nested sampling algorithm. The marginalized posterior distributions obtained from these samples are statistically identical to those obtained by brute force calculations. We find that such Bayesian posteriors can be determined within a few minutes of detecting such transient compact binary sources, thereby improving the chances of their prompt follow-up observations with telescopes at different wavelengths. It may be possible to apply the blueprint of the meshfree technique presented in this study to Bayesian inference problems in other domains.

3.1 Introduction

This chapter is based on the publication *Fast likelihood evaluation using meshfree approximations for reconstructing compact binary sources*, Phys. Rev. D **108**, 064055 (2023).

The detection of gravitational waves (GW) from the GW170817 [127] binary neutron star (BNS) system, followed by the prompt multi-wavelength (gamma-rays to radio) observation of its electromagnetic (EM) counterpart, has led to several fundamental discoveries; and is hailed as a significant breakthrough in astronomy. These discoveries include the validation of long-held hypotheses that BNS mergers are ideal sites for r-process nucleosynthesis and produce short gamma-ray bursts, the first GW-based constraints on the equation of state of nuclear matter in such stars, and the measurement of Hubble constant independent of the cosmic distance ladder.

The inevitable improvement of the detectors' sensitivity in future observation runs is likely to have a two-fold impact on the prospects of multi-messenger observations: *firstly*, the increased bandwidth of the detectors (especially improved sensitivity at low frequencies) will result in a tremendous increase in the computational cost of Bayesian inference of source parameters, including sky localization essential for prompt observation of EM counterparts. Although the `BAYESTAR` [128] algorithm could be used to produce rapid sky maps, it has been recently shown [129] that coherent parameter estimation (PE) can localize the sources better by an average reduction of 14 deg^2 in the uncertainty, underlining the importance of developing fast PE algorithms. *Second*, the reach of the terrestrial network of GW detectors will extend out to several Gpc to the effect that one would have far too many detections of BNS/NSBH signals to contend with whilst generating prompt sky-location maps [46]; so much so that one may have to prioritize the GW sources for EM follow-up based on prospects of new science from a rapid estimation of their mass and spin components as shown by Margalit & Metzger [130], thereby helping EM observatories to use resources optimally. Several fast PE algorithms have been developed recently, such as the coherent multi-detector extension of the relative binning/heterodyne method by Finstaad and Brown (2020) [129], which produces the posterior within twenty minutes for BNS systems with 32 CPU cores. Well-trained machine learning PE methods [131, 132, 133] can significantly reduce the runtimes and produce the posteriors in nearly real-time. In the past, algorithms for accelerated parameter estimation have mainly focussed on speeding up the overlap integral. These include reduced-order models (ROMs) [112, 134, 135], machine-learning aided ROMs [136], Gaussian process regression based interpolation [137] and relative binning [138, 139, 129] algorithms. Our approach takes inspiration from the grid-based likelihood interpolation method [140] based on orthonormal Chebyshev polynomials. The grid-based techniques have a drawback in that the number of interpolation nodes grows exponentially with the dimensionality of the parameter space.

In this work, we propose a new and alternative approach to grid-based likelihood interpolation method [140], a computationally efficient method for evaluating the likelihood function (a key ingredient in Bayesian inference) using meshfree interpolation methods with dimension reduction techniques. We directly interpolate the likelihood function over the parameter space, bypassing the generation of templates and brute-force computation of the overlap integral altogether. Our scheme can quickly approximate the log-likelihood function with high accuracy and produce statistically indistinguishable posteriors over source parameters. Further, both the `GstLAL` search framework [141] and the meshfree method use the idea of dimension reduction using SVD [142], it may be prudent to incorporate this method with the low-latency `GstLAL` search pipeline for

rapid, automated follow-ups of the detected events.

3.2 Bayesian inference

Given data $\mathbf{d} = \mathbf{h}(\vec{\Lambda}_{\text{true}}) + \mathbf{n}$ recorded at a detector containing an astrophysical GW signal $\mathbf{h}(\vec{\Lambda}_{\text{true}})$ embedded in additive Gaussian noise \mathbf{n} , one is interested in solving the inverse problem to estimate the source parameters. Bayesian inference is a stochastic inversion method where the posterior probability density $p(\vec{\Lambda} | \mathbf{d})$ over the source parameters is related to the likelihood function $\mathcal{L}(\mathbf{d} | \vec{\Lambda})$ of observing the data through the Bayes' theorem:

$$p(\vec{\Lambda} | \mathbf{d}) = \frac{\mathcal{L}(\mathbf{d} | \vec{\Lambda}) p(\vec{\Lambda})}{p(\mathbf{d})}, \quad (3.1)$$

where $p(\vec{\Lambda})$ is the prior distribution over the model parameters $\vec{\Lambda} \equiv \{\vec{\lambda}^{\text{ext}}, \vec{\lambda}\}$. In our notation, $\vec{\lambda}$ denotes the intrinsic parameters such as component masses and spins. The set of extrinsic parameters is denoted by $\vec{\lambda}^{\text{ext}}$. We are particularly interested in estimating the extrinsic parameter t_c denoting the fiducial time of coalescence of the two masses. t_c will be mentioned explicitly wherever required, as it is treated in a special way in our analysis.

The forward generative frequency-domain restricted waveform model for non-precessing compact binaries can be expressed as $\mathbf{h}(\vec{\Lambda}) = \mathcal{A} h_+(f_k; \vec{\lambda})$, where the complex amplitude \mathcal{A} depends only on the extrinsic parameters and $h_+(f_k; \vec{\lambda})$ is the '+' polarization of the signal that depends only on the intrinsic parameters [143]. Here $\{f_k\}_{k=0}^{N_s/2}$ defines positive Fourier frequencies, and N_s ¹ is the number of sample points. A GW signal, observed by an interferometric detector, can be considered as a linear combination of the two polarizations weighted by the antenna pattern function. The $h_+(f_k; \vec{\lambda})$ polarization is related to the $h_\times(f_k; \vec{\lambda})$ polarization for non-precessing GW signal as: $h_+(f_k; \vec{\lambda}) \propto i h_\times(f_k; \vec{\lambda})$ [144]. This allows us to write the detector response in terms of any one of the polarizations alone (we have chosen the $h_+(f_k; \vec{\lambda})$ polarization). Using this model, the posterior $p(\vec{\Lambda} | \mathbf{d})$ can be directly evaluated at every point in $\vec{\Lambda}$ using Eq. (3.1). However, in view of the high-dimensionality of $\vec{\Lambda}$, it is more efficient to sample the posterior using stochastic sampling algorithms such as Nested-Sampling [71], or Markov Chain Monte Carlo (MCMC) [143]. From Eq. (3.1), it is evident that for a quick estimation of the posterior distribution, it is imperative to rapidly evaluate the likelihood function.

We work with the phase-marginalized log-likelihood function [145]:

$$\ln \mathcal{L}(\vec{\Lambda}, t_c) = \ln I_0 \left[|\mathcal{A}| z(\vec{\lambda}, t_c) \right] - \frac{1}{2} \|\mathbf{h}(\vec{\Lambda})\|_2^2 \quad (3.2)$$

where $I_0(\cdot)$ is the 0-th order modified Bessel function of the first kind, and $z(\vec{\lambda}, t_c)$ is the frequency-domain overlap-integral:

$$z(\vec{\lambda}, t_c) = 4 \Delta f \left| \sum_{k=0}^{N_s/2} \frac{d^*(f_k) h_+(f_k, \vec{\lambda})}{S_h(f_k)} e^{-2\pi i f_k t_c} \right|, \quad (3.3)$$

¹ N_s = signal duration \times sampling frequency

inversely by $S_h(f_k)$, the detector’s one-sided noise power spectral density (PSD). The data and template vectors are sampled at discrete frequencies $\{f_k\}_{k=0}^{N_s/2}$.

The complexity of evaluating the overlap integral scales directly with the number of data samples, which in turn, scales with the seismic cut-off frequency (approximately) as $N_s \sim f_{\text{low}}^{-8/3}$. As we progress from the O4 observational run ($f_{\text{low}} = 20$ Hz) to O5 at design sensitivity ($f_{\text{low}} = 10$ Hz), evaluating $p(\vec{\Lambda} | \mathbf{d})$ is likely to take at least $\times 6.3$ longer. In addition, additional costs will be incurred in constructing longer templates at the proposal points. Therefore, the likelihood calculation can be expensive. However, our method is immune to this issue as our scheme directly approximates the likelihoods at different sample points. In this work, we have used non-orthonormal radial basis functions (RBF) (Gaussian kernels) centered at interpolation nodes that can be randomly scattered over the volume of the intrinsic parameter space. In this manner, we have effective control of their number in higher dimensional parameter spaces. For completeness, we will begin by exploring the grid-based likelihood interpolation scheme initially introduced by Smith et al. [87, 140], and subsequently, we will delve into the mesh-free interpolation scheme introduced in this thesis.

3.3 Likelihood interpolation

The computational cost of Bayesian inference comprises two parts: the first is incurred in waveform generation (see Appendix 3.7.2), followed by likelihood evaluation at a point proposed by the sampler. Typically, a sampler proposes a large number ($\sim 10^6 - 10^7$) of points to adequately capture the posterior distribution - which makes this part computationally expensive. The other part of the total computational cost can be attributed to the overheads of the sampling method itself. Since the latter cost depends on the efficiency of the sampling algorithm used (and its software implementation) and is significantly less in comparison to the overall cost of PE, we shall ignore it in our discussions. Using likelihood interpolation, we simply bypass the generation of template waveforms and directly interpolate the likelihood over the intrinsic parameter space, thereby reducing the computational cost of likelihood calculation drastically. Both grid-based and meshfree interpolation schemes consist of two stages. In the first stage (start-up stage), we generate interpolating functions for the likelihood while in the second stage (online stage), we rapidly calculate the likelihood, which can be interfaced with a stochastic sampler to estimate the posterior distribution over the sampling parameters. We assume that the parameter estimation is “seeded” by the most significant trigger $\vec{\Lambda}^*$ from an upstream detection pipeline [146, 15]. For an injected signal, this can simply be the injection parameters. The sampling algorithm draws new proposals from a sample space, which is taken to be a moderate-sized “hyper-rectangle” in $\vec{\Lambda}$, centered around the most significant search trigger.

From Eq. (3.2) and (3.3), it is evident that we need to interpolate two pieces: $z(\vec{\lambda}, t_c)$ and $\sigma^2(\vec{\lambda}) \equiv \|h_+(\vec{\lambda})\|_2^2$; and combine them with the amplitude \mathcal{A} to calculate the log-likelihood ratio at a given ‘query’ point $\vec{\lambda}^q$ and a particularly given value of t_c . We will first discuss likelihood interpolation using a grid-based interpolation scheme, which makes use of Chebyshev polynomials to interpolate the likelihood. Then, we will discuss the meshfree interpolation scheme, followed

by the application of these two methods on simulated BBH and BNS systems.

3.3.1 Grid-based likelihood interpolation

Start-up stage

As previously mentioned, during the start-up stage, we construct likelihood interpolants approximating the true likelihood. In the case of a grid-based interpolation scheme, the initial step involves establishing a grid of nodes within the sample space (intrinsic parameter space). Since we are using Chebyshev polynomials (see Appendix 3.7.4) for interpolation, a prudent choice for nodes would be "Chebyshev nodes" known for mitigating Runge's phenomenon. The physical nodes are selected from the sample space and are scaled so that each dimension spans the interval $[-1, 1]$, positioned at the J_{\max} -th order Chebyshev nodes. The locations of these nodes for the i -th dimension are expressed as follows:

$$\lambda_i = \cos\left(\pi \frac{i + 1/2}{J_{\max} + 1}\right), \quad (3.4)$$

where i ranges from 0 to J_{\max} . It can be straightforwardly extended to higher dimensions with each component of $\vec{\lambda}$ following Eq. (3.4). As evident from the Fig. 3.1a, the template norm $\sigma^2(\vec{\lambda})$ is a smoothly varying scalar field over $\vec{\lambda}$. Here $\vec{\lambda}$ consists of a tuple of values $\{\lambda_i\}$ where $i = 1, 2, \dots$ representing each dimensions.

We can interpolate σ^2 at a query point $\vec{\lambda}^q$ by first evaluating the values explicitly at the interpolation nodes $\vec{\lambda}^\alpha$, and then expressing $\sigma^2(\vec{\lambda}^q)$ at an arbitrary point $\vec{\lambda}^q$ as a linear combination of Chebyshev polynomials evaluated at these nodes.

$$\sigma^2(\vec{\lambda}^q) = \sum_{L=0}^{L_{\max}} \sum_{M=0}^{M_{\max}} P_{LM} T_L(\lambda_i^q) T_M(\lambda_j^q) \quad (3.5)$$

where P_{LM} are the projections of the σ^2 onto the Chebyshev polynomials and are given by

$$P_{LM} = \sum_{l=0}^{L_{\max}} \sum_{m=0}^{M_{\max}} T_L(\lambda_l^\alpha) T_M(\lambda_m^\alpha) \sigma^2(\lambda_l^\alpha, \lambda_m^\alpha). \quad (3.6)$$

The unknown coefficients P_{LM} of this linear combination can be uniquely found by enforcing the interpolation criteria as σ^2 are known at N interpolating nodes where $N = \prod_{i=1}^d N_i$ with N_i representing nodes in i -th dimension and d is the total number of dimensions.

On the other hand, as the overlap integral has to be evaluated at an arbitrary point $(\vec{\lambda}^q, t_c)$, it will turn out to be more convenient to interpolate it as a vector. In this case, a set of overlap-integral vectors \vec{z}_α are first constructed at the interpolation nodes, sampled on a uniform grid over t_c . The vectors \vec{z}_α have elements $z_\alpha[k] \equiv z(\vec{\lambda}^\alpha, k \Delta t)$, where Δt is the sampling interval, k is an integer $\in \text{int}([t_c^* \pm \tau]/\Delta t)$, t_c^* represents the 'reference' coalescence time as triggered by the search pipeline and 2τ is the dimension of the sample-space (hyper-rectangle in $\vec{\Lambda}$) along the t_c direction. Since the bulk of the support for the posterior distribution comes from near the peak of these time series, we choose samples that are centered around the triggered value. From Eq. (3.3), it is clear that \vec{z}_α 's can be constructed efficiently using FFT correlations. Once the set

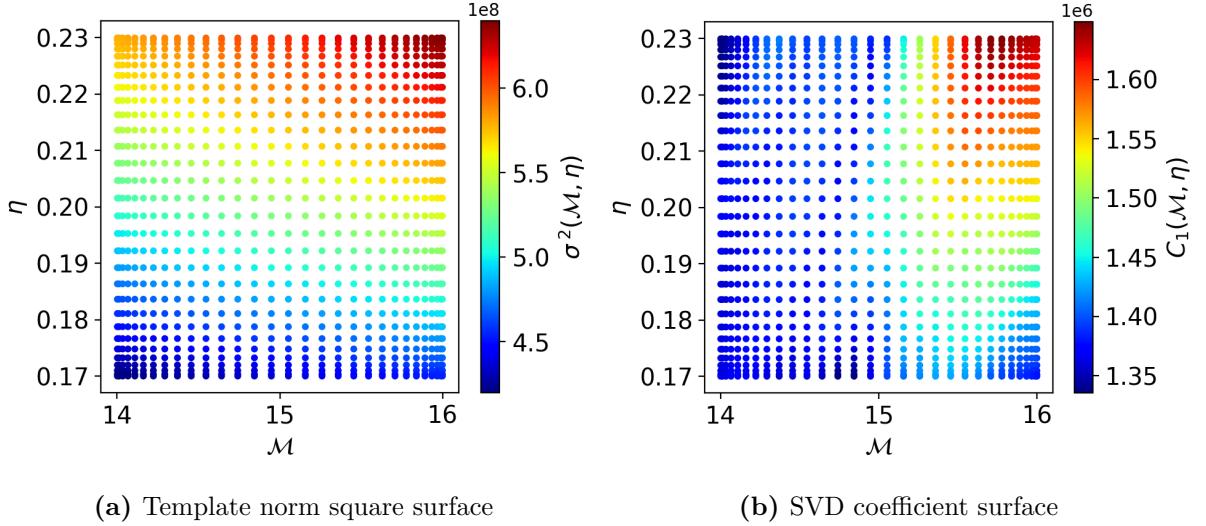


Figure 3.1

of vectors $\{\vec{z}_\alpha\}$ is available, they can be projected over a suitable set of basis vectors $\{\vec{u}_\mu\}$ with linear coefficients $\{C_\mu(\vec{\lambda}^\alpha)\}$ as the following:

$$\vec{z}_\alpha = \sum_{\mu=1}^N C_\mu^\alpha \vec{u}_\mu, \quad (3.7)$$

where C_μ^α is the value of the SVD coefficient at $\vec{\lambda}^\alpha$ for the μ -th basis vector \vec{u}_μ . These basis vectors \vec{u}_μ can be obtained by taking a singular value decomposition (SVD) of the matrix obtained after row-wise stacking of \vec{z}_α . These SVD coefficients appear in the above equation in the decreasing order of their relative importance as determined from the spectrum of singular values. Since there is a strong correlation between \vec{z}_α , the overlap vectors lie in the span of the top- ℓ basis vectors ($\ell \ll N$), which implies that a vector \vec{z}_q at a query point in the sample space can also be spanned by the same set of basis vectors (See Fig. 3.1). As shown in Fig. 3.1b, the SVD coefficients C_μ^α are smoothly varying scalar fields over $\vec{\lambda}$ (sampled at the interpolation nodes), we can use Chebyshev polynomial to interpolate SVD coefficient corresponding to each basis vector \vec{u}_μ as the following:

$$C_\mu^q = \sum_{L=0}^{L_{\max}} \sum_{M=0}^{M_{\max}} P_{LM\mu} T_L(\lambda_i^q) T_M(\lambda_j^q) \quad (3.8)$$

where coefficients $P_{LM\mu}$ are given by

$$P_{LM\mu} = \sum_{l=0}^{L_{\max}} \sum_{m=0}^{M_{\max}} T_L(\lambda_l^\alpha) T_M(\lambda_m^\alpha) C_\mu(\lambda_l^\alpha, \lambda_m^\alpha). \quad (3.9)$$

Now, these coefficients can be found by again imposing the interpolation criteria as they are known at N interpolating nodes. Once we have the interpolants for both $\sigma^2(\vec{\lambda}^q)$ and C_μ^q , we can evaluate them at the query points proposed by a sampler in the online stage.

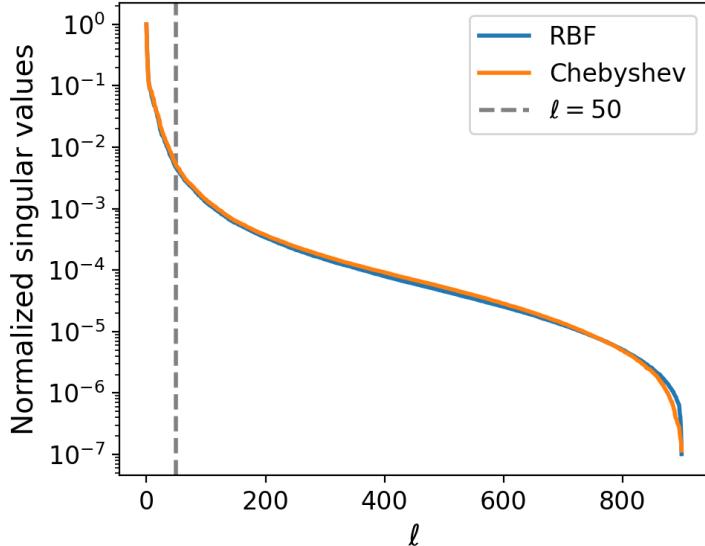


Figure 3.2: Normalized singular value profiles for both Chebyshev and Meshfree interpolation schemes. In both cases, $N = 900$ nodes are chosen where Chebyshev nodes lie on a grid, and meshfree nodes are uniformly scattered across the sample space. The boundaries for the interpolating parameters are also the same for both cases.

Online-stage

In the online stage, a stochastic sampler proposes the query points $\vec{\lambda}^q$ at which the likelihood is evaluated. In a standard (bruteforce) calculation of the likelihood, first, the template waveform is generated at the query point $\vec{\lambda}^q$ and then overlap integral, and template norm square are calculated. In the interpolation scheme, we have the interpolating functions which we are required to be evaluated at the query point $\vec{\lambda}^q$, which are significantly cheaper to evaluate in comparison to bruteforce calculation. Then we combine the C_μ^q with the basis vectors $\vec{\mu}$ in accordance with Eq. (3.7) noting that only the first few basis vectors are required to reconstruct \vec{z}_α within a small error. Note that the interpolating time-series at $\vec{\lambda}^q$ is uniformly sampled over the interval $[t_c^* \pm \tau]$. As such, it is possible that the query t_c does not coincide with the discrete-time samples of \vec{z}_q . In such a case, we use a one-dimensional cubic-spline interpolation to evaluate $z(\vec{\lambda}^q, t_c)$ using a few ‘nearby’ grid samples of \vec{z}_q as input. This implies that \vec{z}_q has to be reconstituted only at a few (~ 10) consecutive sample points, which considerably accelerates the matrix-vector multiplication in Eq. (3.7).

The interpolant for the square of the template norm can be directly evaluated to get the interpolated value $\sigma^2(\vec{\lambda}^q)$. Similarly, the ℓ interpolants for SVD coefficients (Eq. (3.10)) are evaluated to get a set of interpolated coefficients, which are then combined with the corresponding top- ℓ basis vectors \vec{u}_μ (see Eq. (3.7)) to obtain the interpolated overlap-integral \vec{z}_q . Combining the interpolated values $z(\vec{\lambda}^q, t_c)$ and $\sigma^2(\vec{\lambda}^q)$ with the extrinsic-parameter dependent complex amplitude \mathcal{A} (see Eq. (3.2)), we finally obtain the log-likelihood ratio $\ln \mathcal{L}$ at the arbitrary point $(\vec{\lambda}^q, t_c)$. Now, let us discuss the meshfree interpolation scheme.

3.3.2 Meshfree likelihood interpolation

Similar to the grid-based interpolation scheme, the meshfree scheme can also be divided into two stages: (i) a preparatory, start-up stage where we explicitly determine the RBF interpolating functions (interpolants) from the pre-computed likelihood values at the interpolation nodes and (ii) an online stage where these interpolants are evaluated on the fly to ‘predict’ the likelihood values at arbitrary query points in the sample space. Unlike the grid-based scheme (see Fig. 3.3), the nodes are not laid on a specific grid. Instead, for the meshfree scheme, the nodes are randomly distributed around the most prominent trigger identified by the detection pipeline. The subsequent sections will elaborate on the two stages integral to this approach.

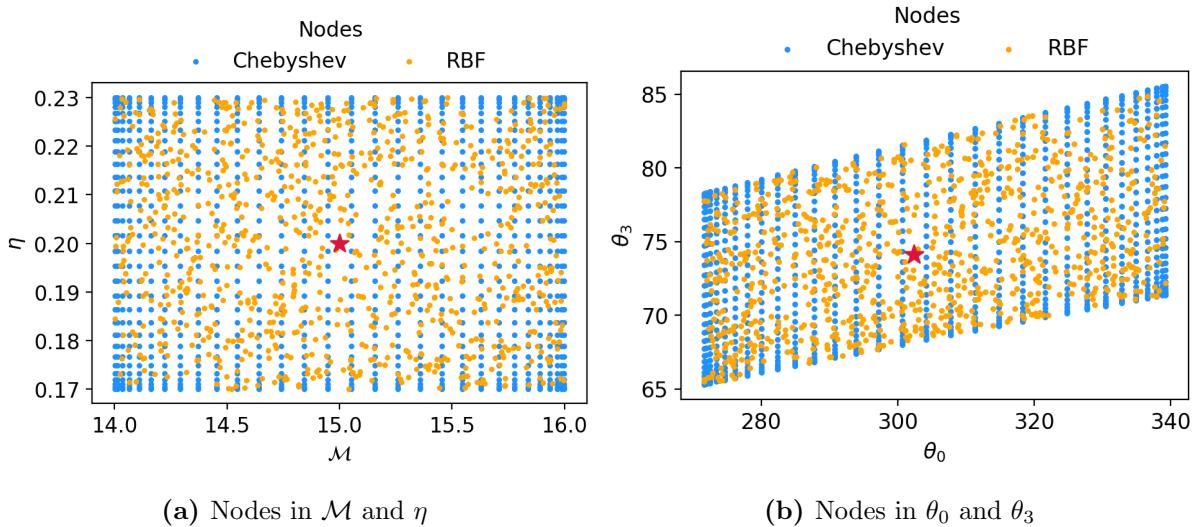


Figure 3.3: Chebyshev and meshfree nodes in the sample parameter space. The red star indicates the BBH injection. For Chebyshev interpolation, we use nodes in \mathcal{M} and η coordinates while for meshfree interpolation, we use θ_0 and θ_3 coordinates as our nodes since the metric varies slowly in θ_0 and θ_3 coordinates in comparison to \mathcal{M} and η .

Start-up stage

In this stage, the nodes are randomly sprayed over the sample space, and the meshfree interpolants are constructed. Similar to the grid-based scheme, we find a suitable set of basis vectors that span the space of N input overlap vectors $\{\vec{z}_\alpha\}$ by stacking these vectors row-wise and performing the SVD of the resultant matrix as shown in the Eq. (3.7). Again, a strong correlation between the \vec{z}_α ’s implies that the overlap vectors lie in the span of the top- ℓ basis vectors ($\ell \ll n$). A vector \vec{z}_q at a query point in the sample space can also be spanned by the same set of basis vectors. Note that for a fixed index μ , the coefficients C_μ^α represent a surface whose values are known only at the input nodes. Along with Eq. (3.7), this implies that the interpolation of the inner-product vector at an arbitrary query point essentially boils down to interpolating the value of the coefficients $C_\mu^q \equiv C_\mu(\vec{\lambda}^q)$. The meshfree interpolants for the SVD coefficient corresponding to the μ^{th} basis vector can be expressed [107] as a linear combination of RBF kernels centered

on the scattered, distinct nodes $\vec{\lambda}^\alpha$, augmented by monomials ranging up-to a specific order:

$$C_\mu^q = \sum_{\alpha=1}^n r_\alpha \phi(\|\vec{\lambda}^q - \vec{\lambda}^\alpha\|_2) + \sum_{j=1}^M b_j f_j(\vec{\lambda}^q), \quad (3.10)$$

where ϕ is the Gaussian kernel (see the Appendix 3.7.3 for various RBF kernels and their corresponding effect on the likelihood reconstruction accuracy) centered on $\vec{\lambda}^\alpha \in \mathbb{R}^d$, and $\{f_j\}$'s are monomials that span the space of polynomials of some preset target degree ν in d variables. In this analysis, we choose $\phi = \exp(-\epsilon \|\vec{\lambda}^q - \vec{\lambda}^\alpha\|_2)$ where ϵ is the shape parameter which determines the extent of the Gaussian RBF. Also, $\mathbf{r} = [r_1, r_2, \dots, r_n]^T$ and $\mathbf{b} = [b_1, b_2, \dots, b_M]^T$ are the set of $(n + M)$ coefficients that need to be uniquely determined to determine the interpolant. Since C_μ^q are known at the interpolation nodes, it allows us to enforce n interpolation conditions. M additional conditions $\sum_{k=1}^n r_k f_j(\vec{\lambda}^k) = 0$, $j = 1, \dots, M$ are added to ensure a unique solution. Together, these lead to a system of equations:

$$\begin{bmatrix} \mathbf{K} & \mathbf{F} \\ \mathbf{F}^T & \mathbf{O} \end{bmatrix} \begin{bmatrix} \mathbf{r} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} C_\mu^\alpha \\ \mathbf{0} \end{bmatrix} \quad (3.11)$$

where the matrices \mathbf{K} and \mathbf{F} have components $K_{ij} = \phi(\|\vec{\lambda}^i - \vec{\lambda}^j\|_2)$ and $F_{ij} = f_j(\vec{\lambda}^i)$ respectively; $\mathbf{O}_{M \times M}$ is a zero-matrix and $\mathbf{0}_{M \times 1}$ is a zero-vector. Eq. (3.11) can be solved uniquely for the unknown coefficients \mathbf{r} and \mathbf{b} , thus completely determining the meshfree interpolant in Eq. (3.10). The solution for \mathbf{r} and \mathbf{b} can be shown to be unique if \mathbf{F} has full column rank. A minimum set of $n = \binom{\nu+d}{\nu}$ interpolation nodes are required to be uniformly distributed over the d -dimensional intrinsic parameter space to determine the coefficients uniquely. Euclidean distances between pairs of points seem to work well, possibly due to the small volume of the sample space in a typical PE analysis. Parameter-space metric-based distances could also be used. A similar procedure is followed to create a separate meshfree interpolant for $\sigma^2(\vec{\lambda}^q)$. Note that by construction, the meshfree interpolation eliminates both the factors that contribute to the high computational cost, namely, (a) generating the forward signal model and (b) explicitly calculating the overlap integral at every query point proposed by the sampling algorithm.

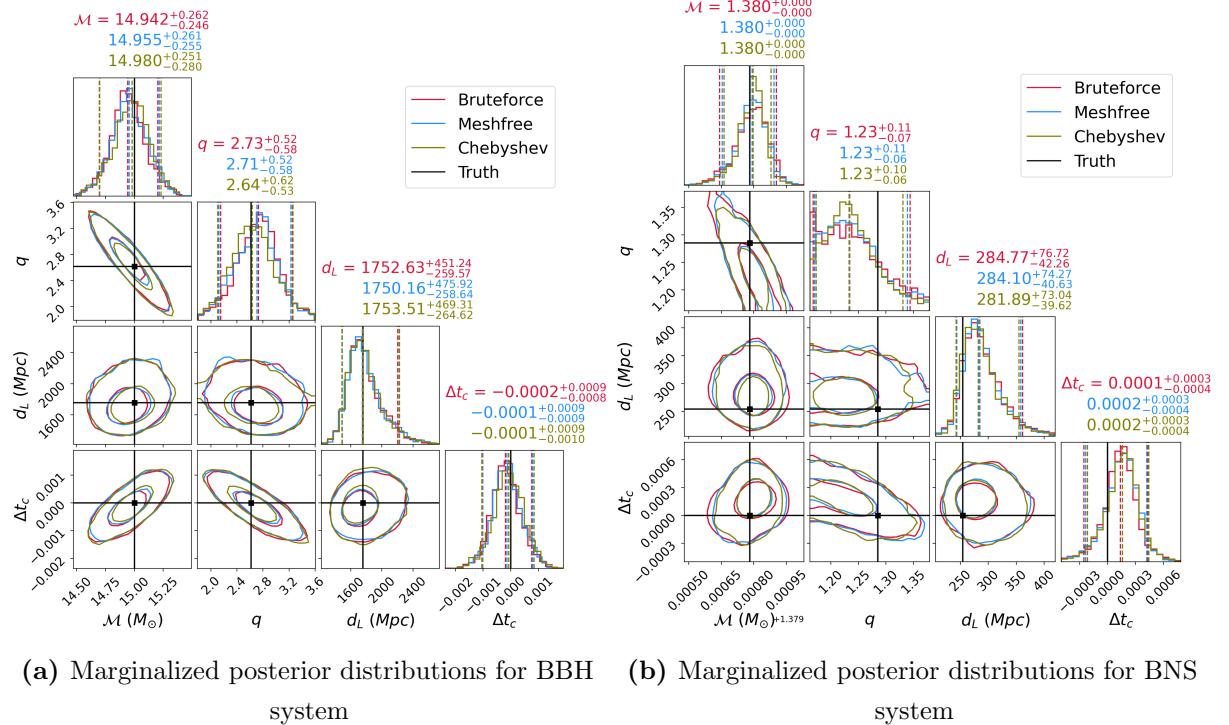
Online stage

In this stage, the $(\ell + 1)$ interpolants (prepared in the offline stage earlier) are evaluated on the fly at arbitrary query points $(\vec{\lambda}^q, t_c)$ proposed by the sampling algorithm. The interpolant for the square of the template norm can be directly evaluated to get the interpolated value $\sigma^2(\vec{\lambda}^q)$. Similarly, the ℓ interpolants for SVD coefficients (Eq. (3.10)) are evaluated to get a set of interpolated coefficients, which are then combined with the corresponding top- ℓ basis vectors \vec{u}_μ (see Eq. (3.7)) to obtain the interpolated overlap-integral \vec{z}_q . As mentioned earlier in the Sec. 3.3.1, the \vec{z}_q is known at the discrete values of times in the interval $[t_c^* \pm \tau]$ and a query t_c may not coincide with those discrete time samples. So we again generate a Cubic spline interpolant of \vec{z}_q centered around query t_c , implying that we only need to reconstruct \vec{z}_q at a few sample times (~ 10) greatly accelerating the matrix multiplication process in Eq. (3.7). Finally, we combine the interpolated values of $z(\vec{\lambda}^q, t_c)$ and $\sigma^2(\vec{\lambda}^q)$ with extrinsic-parameter dependent

complex amplitude \mathcal{A} (see Eq. (3.2)) and obtain the log-likelihood ratio $\ln \mathcal{L}$ at the arbitrary point $(\vec{\Lambda}^q, t_c)$.

3.4 Numerical experiments

3.4.1 2D likelihood interpolation



(a) Marginalized posterior distributions for BBH system (b) Marginalized posterior distributions for BNS system

Figure 3.4: Posterior distributions as obtained by performing PE using bruteforce, Chebyshev, and meshfree likelihood interpolation schemes. Note that the chosen priors over the intrinsic parameter space are narrow which is currently a limitation of our meshfree method. We are currently working on this limitation by employing different strategies to widen the prior widths and will be presented in the future versions of our method.

To demonstrate the accuracy and efficiency of both meshfree and grid-based interpolation methods in reconstructing source parameters, we generated simulated data for two non-spinning compact binary systems: (i) a Binary Black Hole (BBH) system and (ii) a Binary Neutron Star (BNS) system. The details of the injections and the simulated data are provided in the table below. Both signals were injected into the `aLIGOZeroDetHighPower` noise, simulating a configuration of aLIGO design sensitivity for a single detector, namely 'L1,' such that the matched-filter (SNR) is approximately ~ 10 .

We performed Bayesian inference on these simulated data using both (a) *bruteforce* likelihood calculation used in PyCBC inference and (b) and by using both grid-based interpolation and the proposed meshfree likelihood interpolation scheme outlined in earlier sections. We used publicly available software [118] for radial basis functions and the `Dynesty` [72, 73] nested-sampling package for carrying out the Bayesian inference analysis. We varied two intrinsic parameters

	\mathcal{M}/M_{\odot}	η	d_L (Mpc)	t_c	duration (secs)	SNR
BBH	15	0.20	1758	1126258006	8	10
BNS	1.379	0.246	254	1126258250	256	10

Table 3.1: Injection parameters of the simulated GW events. For both cases, we chose $\alpha = \delta = \iota = \pi/4$, and $\phi_c = \psi = 0$. The seismic cutoff frequency was chosen to be 20 Hz, whereas the higher cutoff frequencies were 1024 Hz and 1500 Hz for BBH and BNS systems, respectively.

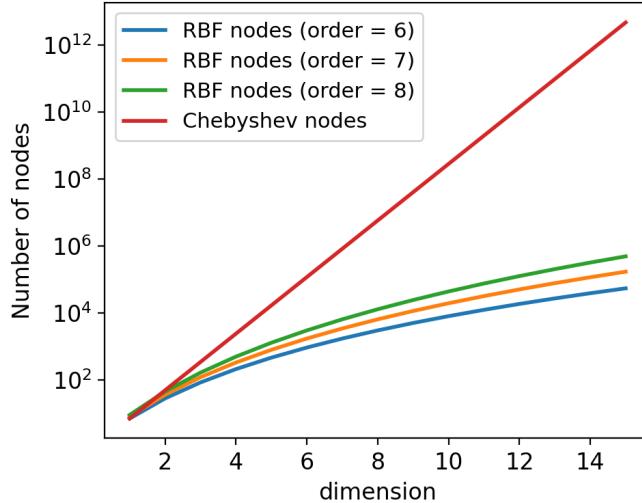


Figure 3.5: Scaling of nodes with the number of dimensions. The number of Chebyshev nodes increases exponentially in stark contrast to those of meshfree nodes.

(component masses) and two extrinsic parameters (luminosity distance and coalescence time), keeping other parameters fixed. For the grid-based scheme, we opted for a Chebyshev grid of dimensions 30×30 in \mathcal{M} and η . In contrast, for the meshfree method, we distributed uniformly random nodes across the same boundaries as in the Chebyshev case, as illustrated in Fig. 3.3. In both approaches, we utilized a total of $N = 900$ interpolating nodes. In the meshfree case, we selected $\phi = \exp(-\epsilon r^2)$ as the RBF kernel and employed θ_0 and θ_3 as interpolating nodes, derived straightforwardly from the corresponding nodes in \mathcal{M} and η coordinates. In both scenarios, the top $\ell = 50$ basis vectors were employed. For the BBH case, we set $\nu = 1$ and 0 for σ^2 and C_{μ}^q interpolants, respectively. For the BNS system, $\nu = 2$ and 0 were chosen for σ^2 and C_{μ}^q interpolants, respectively. Another pivotal parameter influencing the accuracy of the interpolating functions in the meshfree method is ϵ , determining the spread of the Gaussian at each node. In the BBH case, we employed $\epsilon = 0.75$ and 0.5 for σ^2 and C_{μ}^q interpolants, respectively. For the BNS system, $\epsilon = 0.5$ and 0.6 were selected for σ^2 and C_{μ}^q , respectively. We found an absolute error of $\sim 10^{-2}$ across the sample space in approximating the log-likelihood function for both schemes. As depicted in Fig. 3.4, the posterior distributions of source parameters using both Chebyshev and meshfree methods align well with the brute-force method. It's worth noting that for higher dimensions, the grid-based scheme requires an exponentially increasing number of Chebyshev

nodes compared to meshfree nodes, as illustrated in Fig. 3.5. Deploying a grid-based scheme for likelihood interpolation becomes impractical due to the escalating computational expense associated with generating interpolants and evaluating likelihoods. In contrast, the meshfree method, by its design, is independent of any grid and thus suitable for higher-dimensional problems. In the subsequent subsection, we illustrate an example involving a simulated BNS system and estimate its source parameters using our proposed meshfree method, where the dimensionality of the interpolating parameters is four.

3.4.2 4D likelihood interpolation

To demonstrate the accuracy and speed of the meshfree method in reconstructing the source parameters, we prepared synthetic 360 seconds long data \mathbf{d} , sampled at 4096 Hz. For this, a simulated GW signal h from a canonical BNS system with component masses $m_{1,2} = 1.4 M_\odot$ and mass-weighted effective dimensionless spin $\chi_{\text{eff}} = 0.05$ was injected (using the `IMRPhenomD` [102] signal model) in colored Gaussian noise using the noise power spectral density model [115] of `aLIGO` detectors. The distance to the source was adjusted for a moderate matched-filtering SNR of 10. The seismic cut-off frequency was chosen to be 20 Hz to mimic data from the upcoming O4 science run.

	\mathcal{M}/M_\odot	η	χ_{eff}	SNR
Injection	1.2187	0.25	0.05	10.00
Standard	1.2187 1.2185	0.249 0.246	0.250 0.246	0.0507 0.0493
Meshfree	1.2187 1.2185	0.249 0.246	0.250 0.246	0.0507 0.0493

Table 3.2: Reconstruction of a canonical BNS event.

We performed Bayesian inference on this simulated data using both (a) *direct* likelihood calculation used in PyCBC inference and (b) and by using the proposed meshfree likelihood interpolation scheme outlined in earlier sections. We used publicly available software [118] for radial basis functions and the `Dynesty` [72, 73] nested-sampling package for carrying out the Bayesian inference analysis. We varied four intrinsic parameters (component masses and aligned-spin magnitudes) and two extrinsic parameters (luminosity distance and coalescence time), keeping other parameters fixed.

For this exercise, we used $n = 800$ random input nodes over $\vec{\lambda}$. The top $\ell = 120$ basis vectors and a polynomial order $\nu = 6$ with a corresponding nominal median relative error $\sim 10^{-5}$ across the sample space in approximating the log-likelihood function. The accuracy trade-offs of likelihood reconstruction by varying basis size is shown in the Appendix 3.7.1.

The meshfree parameter reconstruction was completed in 5.3 min in comparison to 31.7 h taken by the direct calculation. The likelihood function was evaluated 676 times faster using the meshfree method. Some of the estimated parameters have been compared in Table 3.2, which show identical values obtained by both methods. The marginalized PDF over five parameters \mathcal{M} ,

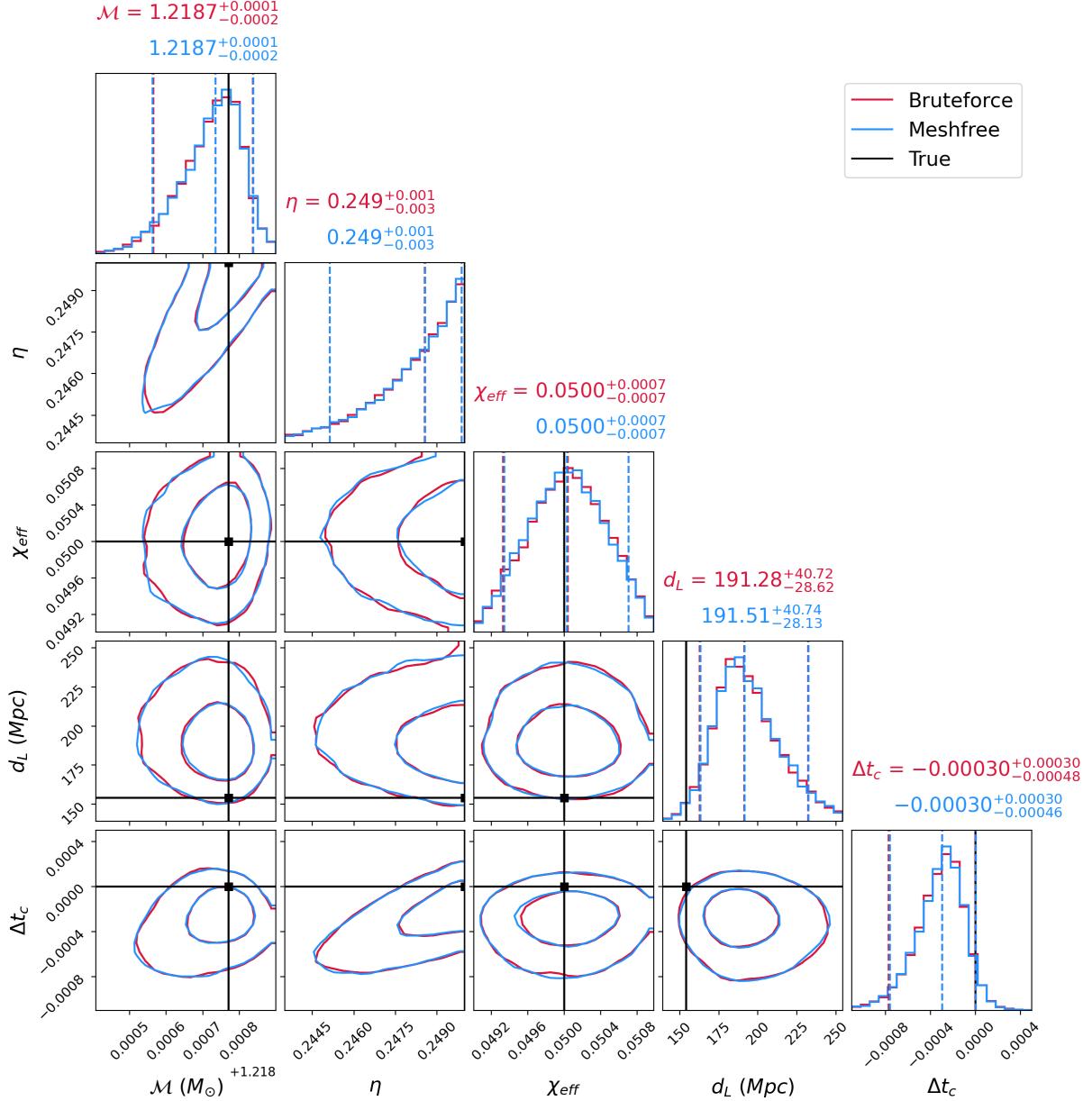


Figure 3.6: Marginalized PDF for chirp mass \mathcal{M} , symmetric mass ratio η , and effective spin χ_{eff} , d_L , and Δt_c parameters of a simulated BNS event at a seismic cutoff of 20 Hz. The injection parameters are shown as black lines. The 50% and 90% contours for meshfree method (blue) and Bruteforce (red) are also shown. The plot-overlaid marginalized PDF obtained from the proposed meshfree method (blue) and Bruteforce (red) is virtually indistinguishable.

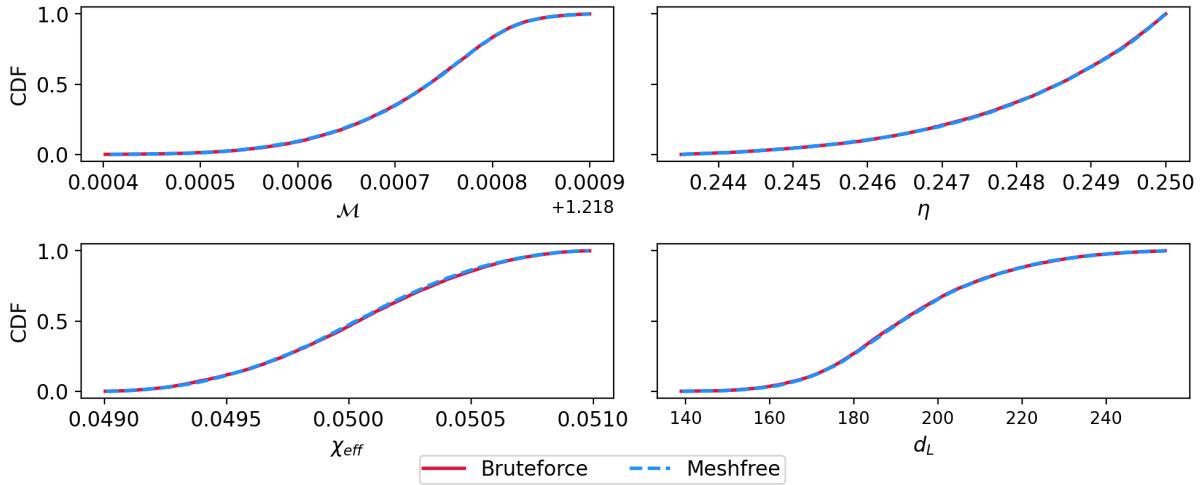


Figure 3.7: The cumulative distribution functions (CDFs) of \mathcal{M} , η , and χ_{eff} , d_L using both meshfree (dashed blue) and bruteforce (red) method are shown and are in good agreement with each other.

η , χ_{eff} , d_L , and Δt_c are shown in the corner plot Fig. 3.6. The Fig. 3.7 shows cumulative density (CDF) profiles of these distributions obtained from both Bruteforce and the meshfree method plotted together. While both the PDF and CDF profiles look virtually indistinguishable, we also calculate statistical measures of similarity [147] using the PDFs obtained from the two methods for further validation: both the Kolmogorov-Smirnov statistic (0.0130) and the Bhattacharyya distance (0.0006) between the chirp-mass PDF profiles support the fact that the two distributions are nearly identical. We get similar results for posterior distributions of other parameters.

This numerical example shows that the meshfree method can generate a statistically indistinguishable replica of the posterior distributions in a GW Bayesian inference problem at a small fraction of the total computational cost.

3.5 Speed-up analysis

We calculated the ratio of (average) time taken to compute the log-likelihood by these techniques at a fixed accuracy of reconstruction. Several simulated data sets were used in this study, generated by injecting signals having different parameters in colored Gaussian noise using the aLIGO noise model at a fixed matched-filtering SNR of 10. Seismic cutoff frequencies at 20 Hz (10 Hz) were considered to mimic data from upcoming O4 (O5) science runs.

There is an obvious trade-off between accuracy and speed-up of the meshfree method, which are determined by the choice of (n, ℓ, ν) parameters. Larger values can lead to more accurate likelihood estimates, albeit at a higher computational cost and vice-versa. We used a heuristic combination to guarantee median (relative) errors $\lesssim 9 \times 10^{-5}$ in the estimated log-likelihood values across the entire sample space. The log-likelihood function was evaluated and timed for a large number of random points uniformly distributed over the $(\vec{\lambda}, t_c)$ space.

Table 3.3 summarizes the speed-ups corresponding to the two different seismic cutoff frequencies for three compact binary systems with equal component masses. It is further elucidated

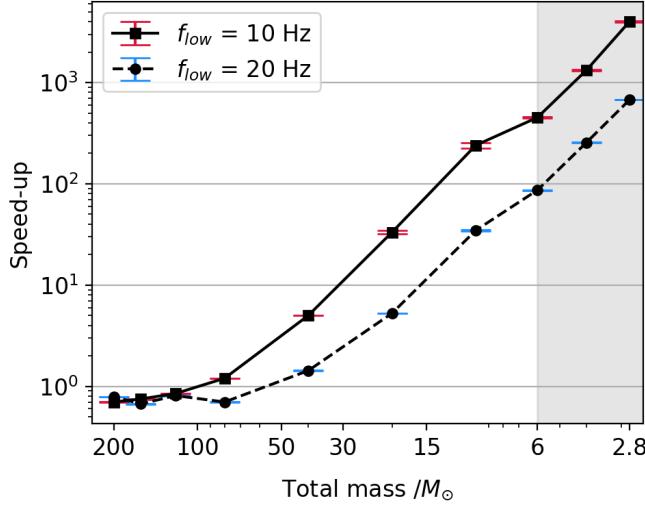


Figure 3.8: Speed-up factors (averaged over several thousand evaluations) in calculating the log-likelihood function with the meshfree method for different equal-mass ($q = 1$) systems with (dimensionless) spin magnitude of 0.05 (0.2) for BNS (BBH) systems. The shaded region denotes BNS systems. The combination of (n, ℓ, ν) parameters were tuned for a median error of $\sim 10^{-5}$.

seismic cut-off	M/M_{\odot}	t_{mf} /ms	t_{pycbc} /ms	speed-up
$f_{low} = 10$ Hz	2.8	0.90	3604.76	4005.2
	4.0	1.06	1405.84	1326.2
	20.0	0.88	29.27	33.2
$f_{low} = 20$ Hz	2.8	0.67	452.93	676.0
	4.0	0.81	207.35	256.0
	20.0	0.63	3.32	5.3

Table 3.3: The first and second columns define the seismic cut-off frequency and total mass of the injected GW signal, respectively. The third and fourth columns show the median time (in ms) taken for a single evaluation of the log-likelihood function using the standard PyCBC method and our approach at a nominal relative error of $\mathcal{O}(10^{-5})$. The last column provides the relative speed-up of our method in comparison with the standard likelihood calculation. Median time (in ms) taken for a single evaluation of the log-likelihood function using the standard PyCBC method and the meshfree approach at a nominal relative error of $\sim 10^{-5}$.

in Fig. 3.8, where the speed-up comparison is drawn between a larger number of equal-mass binary systems covering a wider range of parameters. From Table 3.3 and Fig. 3.8, it is clear that the likelihood computation can be sped up ~ 4000 times faster for canonical BNS systems at a nominal error of $\sim 10^{-5}$ using the meshfree method against standard likelihood implementation in PyCBC inference. However, this would not reflect the true benefit of our proposed method, as PyCBC inference is not optimized to calculate the fast posterior distribution for LIGO. Our method computes the posterior within a few minutes for the BNS system. Thus, our proposed

scheme has the potential to perform rapid reconstruction of source parameters in upcoming observation runs for `aLIGO` detectors similar to the other existing optimized methods. The low-mass systems with a large number of in-band cycles would benefit most from the meshfree method. In contrast to the standard method, the time taken by the meshfree method is relatively unaffected by the chirp time of the signals. All the tests were performed on a single-core AMD EPYC 7542 CPU@2.90GHz CPU.

3.6 Conclusion and Outlook

We have presented an alternative approach to the grid-based method [140], which is computationally efficient for accurately evaluating the log-likelihood function. Our meshfree method can be easily integrated into well-known sampling algorithms (e.g., MCMC, Nested Sampling) to substantially accelerate the Bayesian inference of source parameters of coalescing compact binary sources, triggering prompt observation of their EM counterparts in the future. Using synthetically generated data of a GW signal from a merging BNS system, we have demonstrated that the posterior distributions are statistically identical to those obtained using the standard PyCBC inference [16]. For BNS systems, the likelihood function can be calculated ~ 4000 times faster at any point proposed by the sampling algorithm than the direct likelihood calculation implemented in PyCBC inference. At this point, we want to remind the readers that we made the comparison of our scheme against PyCBC inference only to verify the robustness of our scheme. PyCBC inference is not used for the fast PE run by LVK. Therefore, for a fair speed-up comparison, we must compare our methods against those schemes [129, 131, 132, 112, 134, 135, 137, 138, 139] computing the real-time posterior distribution. However, the comparison with the optimized schemes is beyond the scope of the current work. In the follow-up works, we will compare our method against those schemes in detail. Numerical experiments with a coherent, multi-detector implementation of the meshfree algorithm suggest that we can solve the coherent BNS PE problem (over a ten-dimensional parameter space) in ~ 164 seconds using 64 CPU cores. Further optimizations and runs over the full parameter space are underway. These details are available in a follow-up paper [148]. It may be prudent to incorporate this method with the low-latency `GstLAL` [146] search pipeline for rapid, automated follow-ups of the detected events since both use the idea of dimension reduction using SVD. In the current implementation of this algorithm, we need to create meshfree RBF interpolants from scratch for each new event triggered by the search pipelines. However, this task is embarrassingly parallel and can benefit from multiple CPU cores to expedite the preparatory stage. Future refinements could involve using the Fisher matrix [149, 150] as a guide to identifying the sample space volume and using sophisticated interpolation node distribution algorithms [151]. Finally, the techniques of dimension reduction and meshfree approximation could be applied to situations where the likelihood function varies smoothly over the sample space. It is thus possible to adapt this idea to Bayesian inference in fields as diverse as cosmology, biochemical kinetic processes, and systems biology.

In the next chapter, we discuss the multidetector extension of this method to enable sky-localization of sources along with other angles as well.

Acknowledgements

We thank M. K. Haris for various suggestions and for his help at the conceptual stages of this work. We thank M. Vallisneri and T. Dent for carefully reading the manuscript and for offering several comments and suggestions to improve the presentation and content of the paper. We also thank other LSC colleagues, S. Mitra, A. Ganguly, and S. Kapadia, for their useful comments. L. P. is supported by the Research Scholarship Program of Tata Consultancy Services (TCS). A. R is supported by the research program of the Netherlands Organisation for Scientific Research (NWO). A. S. gratefully acknowledges the generous grant provided by the Department of Science and Technology, India, through the DST-ICPS cluster project funding. We would like to thank the HPC support staff at IIT Gandhinagar for their help and cooperation.

3.7 Appendix

3.7.1 Accuracy trade-off: Effect of varying basis size

We present the results of a simulation to demonstrate the accuracy of the mesh-free method on the basis of size. We inject a GW signal from a compact binary system with component masses ($1.4M_{\odot}$, $1.4M_{\odot}$) in simulated noise generated using the advanced LIGO design sensitivity curve with a lower seismic cut-off frequency of 10 Hz. On the one hand, we compute the log-likelihood-ratio (LLR) using the standard functions available in the PyCBC package (considered to be the ‘ground truth’ for this simulation) and compare it with the approximate value using the mesh-free interpolation method as outlined in our manuscript. The meshfree method uses 10^3 randomly chosen initial nodes over the sample space. We vary the number of the basis vectors for reconstructing the likelihood and estimate the difference between interpolated likelihood and the ‘ground-truth’ likelihood value. For these simulations, we observe that the top 20 basis vectors are sufficient to approximate the likelihood with sufficient accuracy. However, we demonstrated the absolute error between the true and approximated values for 5, 7, 20, 50, 100, and 200 basis vectors, respectively. Fig. 3.9a shows the probability distribution function (PDF) of the difference between true and estimated likelihood, and Fig. 3.9b shows the corresponding cumulative distribution function (CDF). It is clear from these figures that as we increase the number of basis vectors, the error distribution becomes more concentrated around zero as compared to the relatively flattened distribution for smaller basis sizes.

3.7.2 Scaling of the speed-up factor with waveform duration

The meshfree likelihood evaluation has no dependence on the length of the waveform. On the other hand, in the standard method, a significant time is spent on the waveform generation, followed by the evaluation of the likelihood integral. The latter depends on the length of the waveform. As shown in Table II, the speed-up ratios decrease with higher masses (shorter

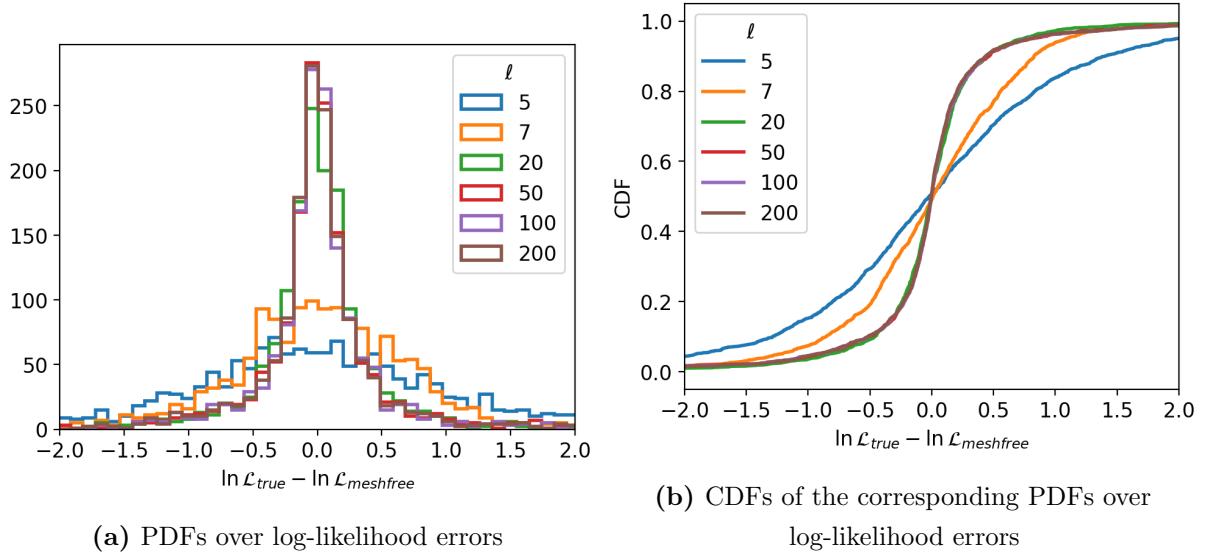


Figure 3.9: Chebyshev and meshfree nodes in the sample parameter space. The red star indicates the BBH injection. For Chebyshev interpolation, we use nodes in \mathcal{M} and η coordinates while for meshfree interpolation, we use θ_0 and θ_3 coordinates as our nodes since the metric varies slowly in θ_0 and θ_3 coordinates in comparison to \mathcal{M} and η .

waveforms). To verify this, we calculated the likelihood evaluation time using both bruteforce

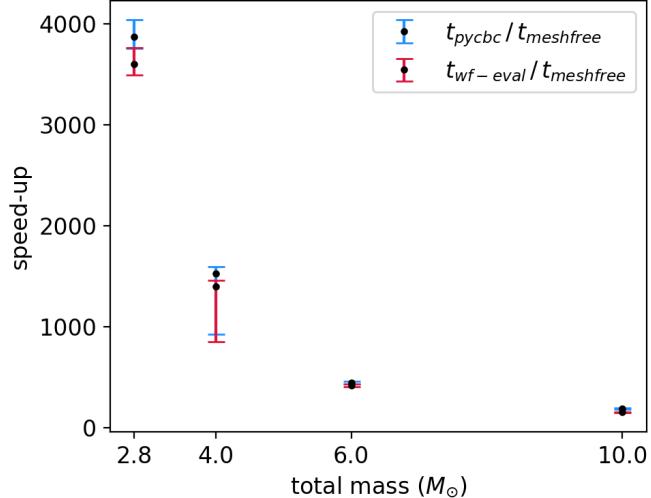


Figure 3.10: The relative speed-up between the likelihood evaluation using PyCBC and the proposed meshfree method is shown (blue). Also, The ratio of overall time for generating waveform generation part (total mass: $2.8M_{\odot} - 10M_{\odot}$) using PyCBC and the likelihood evaluation via meshfree scheme has been compared (red). The dominant cost in the traditional likelihood calculation arises from the generation of long-duration waveforms.

(t_{pycbc}) and meshfree method (t_{rbf}) and also evaluated the waveform generation time ($t_{\text{wf eval}}$) for the four compact binary sources at 10^4 points each. As shown in Fig. 3.10, the waveform generation part is the dominant cost in traditional calculations. For a BNS system, it takes about ~ 7 times longer to generate the waveform as compared to the time for evaluating the likelihood

integral. As we go towards heavy CBC systems, the speed-up ratio decreases as expected.

3.7.3 RBF kernels and meshfree interpolation

In our analysis we chose $\phi = \exp(-\epsilon \|\lambda^q - \lambda^\alpha\|_2)$. For the sake of completeness, we also performed the 2D meshfree likelihood interpolation using three different RBFs apart from the Gaussian RBF kernel (see table 3.4). As shown in the table above, the RBF kernels ϕ also depend on

symbol	ϕ	SVD coeff	σ^2
‘ga’	$\exp[-(\epsilon r)^2]$	0.5	0.75
‘iq’	$[1 + (\epsilon r)^2]^{-1}$	0.25	0.5
‘imq’	$[1 + (\epsilon r)^2]^{-1/2}$	0.35	0.60
‘mq’	$-[1 + (\epsilon r)^2]^{1/2}$	0.25	0.5

Table 3.4: RBF kernels

the shape parameter (ϕ), which affects the interpolation accuracy. The shape parameter can be optimized using the “Leave One Out Cross-Validation” (LOOCV) procedure [117]. In this procedure, we estimate the optimized value of the epsilon (given a range) by constructing the interpolant at each value of ϵ using all the nodes but one at which we evaluate the interpolant and calculate the error at that point, and it is repeated for all the nodes. Then, the rms value of the error is calculated using the errors at all the nodes. So, at each value of ϵ , we have a root-mean-square (RMS) error, and the value of the ϵ giving the minimum RMS error is the optimized value. However, the values shown in the table 3.4 are chosen heuristically. In chapter 2, we find the shape parameter using the LOOCV method as implemented in the RBF python package [118].

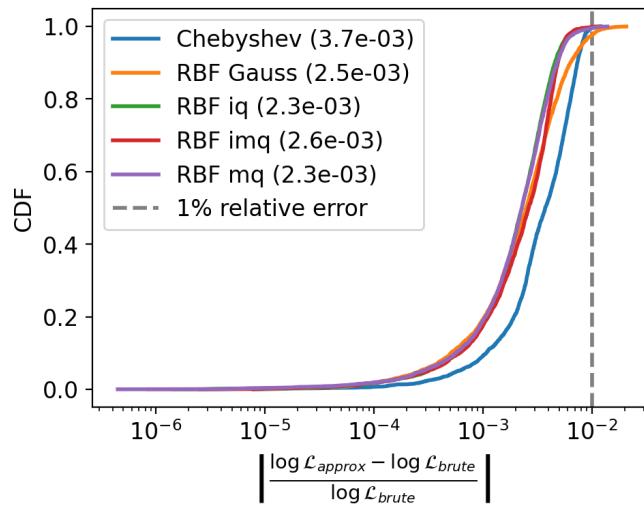


Figure 3.11: Relative error in likelihood reconstruction for Chebyshev interpolation, and meshfree interpolation with different RBF kernels. The median relative errors are shown in the legend, and all are $\sim \mathcal{O}(10^{-3})$ for all the kernels.

3.7.4 Chebyshev polynomials

Chebyshev polynomials for a single dimension are given as

$$T_L(x) = \frac{(x - \sqrt{x^2 - 1})^L + (x + \sqrt{x^2 - 1})^L}{2w} \quad (3.12)$$

where $w = \sqrt{(1 + \delta_{L0})(L_{\max} + 1)/2}$ is a normalization factor for the chebyshev polynomials and δ_{L0} is the Kroenecker delta. These polynomials satisfy the discrete orthogonality condition,

$$\sum_{n=0}^{L_{\max}} T_K(x_n)T_M(x_n) = \delta_{KM} \quad (3.13)$$

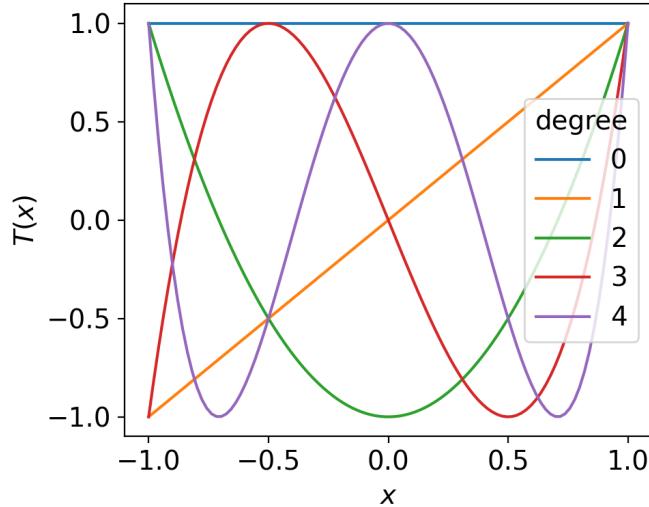


Figure 3.12: Chebyshev polynomials with different degrees

Chapter 4

Prompt sky localization of compact binary sources using meshfree approximation

Abstract The number of gravitational wave signals from the merger of compact binary systems detected in the network of advanced LIGO and Virgo detectors is expected to increase considerably in the upcoming science runs. Once a confident detection is made, it is crucial to reconstruct the source’s properties rapidly, particularly the sky position and chirp mass, to follow up on these transient sources with telescopes operating at different electromagnetic bands for multi-messenger astronomy. In this context, we present a rapid parameter estimation (PE) method aided by mesh-free approximations to accurately reconstruct properties of compact binary sources from data gathered by a network of gravitational wave detectors. This approach builds upon our previous algorithm [L. Pathak *et al.*, Fast likelihood evaluation using meshfree approximations for reconstructing compact binary sources, Phys. Rev. D **108**, 064055 (2023)] to expedite the evaluation of the likelihood function and extend it to enable coherent network PE in a ten-dimensional parameter space, including sky position and polarization angle. Additionally, we propose an optimized interpolation node placement strategy during the start-up stage to enhance the accuracy of the marginalized posterior distributions. With this updated method, we can estimate the properties of binary neutron star (BNS) sources in approximately 2.4 (2.7) minutes for the `TaylorF2` (`IMRPhenomD`) signal model by utilizing 64 CPU cores on a shared memory architecture. Furthermore, our approach can be integrated into existing parameter estimation pipelines, providing a valuable tool for the broader scientific community. We also highlight some areas for improvements to this algorithm in the future, which includes overcoming the limitations due to narrow prior bounds.

4.1 Introduction

This chapter is based on the publication *Prompt sky localization of compact binary coalescences using meshfree approximation*, Phys. Rev. D **109**, 024053 (2024).

Gravitational wave signals emitted by compact binary systems made of neutron stars and black holes are expected to take up to a few minutes to sweep through the sensitive band of the advanced LIGO and Virgo detectors, leading up to the epoch of their eventual cataclysmic merger. The merger of two neutron stars/Neutron star-black holes (NSBH) [152, 153, 154, 155, 156]) can be followed by the emission of electromagnetic (EM) radiation at different wavelengths that fade over different time scales ranging from a few minutes to several months, as seen in the fortuitous detection of the binary neutron star system (BNS) GW170817 [13] by the network of advanced LIGO-Virgo detectors [157, 158, 159, 160, 161] in the second science run. This discovery resulted in the first multi-wavelength observation of post-merger electromagnetic (EM) emissions from the merger of two neutron stars. The successful localization and discovery, in this case, were primarily due to the source’s fortunate proximity: GW170817 was relatively close (about 40 Mpc), falling well within the sky- and orientation-averaged ranges of the two LIGO detectors, resulting in the loudest signal ever detected by a GW network. The discovery of the EM counterpart of GW170817 improved understanding of the physics of short gamma-ray bursts (SGRBs)[162], confirmation of the formation of heavy elements after the merger and other areas of new physics, and provided invaluable opportunities to explore the mysteries of the universe. Therefore, BNS systems are the prime targets for multi-messenger astronomy.

In upcoming observing runs, the LIGO-Virgo detectors [157, 158, 159, 160, 161] are expected to be sensitive above 10 Hz, enabling a more extensive exploration of cosmic volume compared to their current versions. Consequently, the detection rate is expected to increase to the point where signals from compact binary mergers can be observed daily. It is project that there could be $10^{+52}_{-10}/1^{+91}_{-1}$ BNS/NSBH events detected in the O4 run [46]. To efficiently observe EM counterparts in the future, it will be crucial to prioritize events based on their chirp masses, as suggested by Margalit et al. [130]. Furthermore, future GW events could provide opportunities to study various physical properties of binary systems, such as those with eccentricity [163], spin precession [164, 165], and more. As such, prompt localization and source reconstruction would be vital for studying electromagnetic (EM) counterparts of binary neutron star systems (BNS) or neutron star-black hole (NSBH) binary systems in the future. Moreover, a fast PE technique can facilitate the execution of extensive inference investigations [166], like population inference, within a reasonable time frame, which could be impractical when employing conventional brute-force PE methods.

The standard technique for PE uses Bayesian inference, which involves calculating the likelihood of observing the data under the hypothesis that an astrophysical signal (with some model parameters) is embedded in additive Gaussian noise with zero mean and unit variance. This likelihood calculation involves two computationally expensive parts: first, generating the model (template) waveforms [81, 85, 167] at sample points proposed by the sampling algorithm and, thereafter, determining the overlap between the data and the waveform. Generating these templates is the most computationally expensive part of the likelihood calculation, especially for low-mass systems like binary neutron stars (BNS). This computational burden is amplified by the improved sensitivity of detectors, resulting in a larger number of waveform cycles within

the detectors' sensitive band for BNS systems. Additionally, incorporating physical effects into the waveform can further increase the computational cost of generating the waveform and sampling from the expanded parameter space. Wherever possible, factorizing [168] the log-likelihood function into pieces that exclusively depend on the extrinsic and intrinsic parameters, respectively, can lead to a significant reduction in the number of waveform generation and thereby speed-up the PE analysis. For example, such factorizations can be made for non-precessing waveforms, as considered in this work.

In the current operational framework of the LIGO-Virgo-KAGRA (LVK) collaboration, the primary tool for rapid sky localization is **BAYESTAR** [128] - a Bayesian, non-Markov Chain Monte Carlo (MCMC) sky localization algorithm. This method exhibits remarkable speed, furnishing posterior probability density distributions across sky coordinates within a few tens of seconds following the detection of a gravitational wave signal from the merger of a compact binary source. However, through a comprehensive Bayesian PE analysis encompassing both intrinsic and extrinsic model parameters, Finstad et al. [129] have established that the accuracy of sky localization can be significantly augmented, achieving an enhancement of approximately 14 deg^2 over those obtained from the **BAYESTAR** algorithm alone. Beyond the evident advantages of sky-location precision, a full PE analysis also furnishes a highly accurate estimation of the chirp mass for compact binary systems. This additional information plays a pivotal role in making judicious decisions regarding electromagnetic (EM) follow-up observations - which would be crucial in future observing runs of the network of advanced detectors, further underscoring the merits of developing fast-PE algorithms. Deep-learning-based sky-localization tools (CBC-SkyNet [169]) have been recently developed that can obtain “pre-merger” sky localization areas that are comparable in accuracy to **BAYESTAR**.

Numerous rapid PE algorithms have surfaced in the gravitational-wave literature, which revolves around two overarching concepts:

- i. Likelihood-Based Approaches: The first category involves approaches that rapidly evaluate the likelihood function. This set of methods encompasses various methods such as Reduced Order Models (ROMs)[111, 112, 170, 113, 106], Heterodyning (or relative binning) [139, 138, 168], the simple-pe [171] algorithm, and methods based on Gaussian process regression (GPR) [172] like RIFT [137]. Other techniques, such as adaptive frequency resolution-based likelihood evaluation [124] and mass-spin reparametrization-based rapid PE [173], have also been proposed. Improved algorithms allowing for more efficient reduced-order quadrature bases have been recently proposed [114].
- ii. Likelihood “Free” Approaches: The second category consists of methods that bypass direct likelihood evaluation and instead learn the posterior distributions using advanced machine-learning (ML) techniques such as Deep Learning, Normalizing Flows, and Variational Inference [174, 175, 176, 132].

Hybrid techniques that combine likelihood heterodyning with tools to enhance the convergence of gradient-based MCMC samplers have also been proposed [177]. Other techniques involving score-based diffusion models [178] to learn an empirical noise distribution directly from the

detector data have also been proposed. In this context, our mesh-free approach aligns with the first category of rapid PE methods since it is designed to swiftly assess the likelihood. Note that the RIFT [137] method also interpolates the likelihood directly using GPR. In comparison, the meshfree method first represents the likelihood function on an orthonormal basis and then uses radial basis functions to interpolate the coefficients. So, inspite of the apparent similarity, the context in which the “interpolation” technique is applied is completely different in the two methods. The use of GPR (a supervised training method) to “learn” the likelihood function at different points of the parameter space is in the spirit of ML methods in contrast with the classical approach followed by the meshfree method.

In our previous work [179], we introduced the meshfree method as a means of rapidly inferring parameters for compact binary coalescence (CBC) sources for a single detector by considering only six parameters: the two-component masses, two aligned spins, luminosity distance, and coalescence time. In this study, (a) we extend the method to include additional parameters such as sky location, inclination, and polarization, thereby enabling a coherent multi-detector PE analysis. (b) Furthermore, we have modified the Radial Basis Function (RBF) meshfree interpolation node placement scheme, which directly impacts the accuracy of the PE results. In contrast to our previous work, which utilized uniformly distributed nodes over the sample space, we now employ a combination of multivariate Gaussian and uniform distributions for efficient node placement. (c) Our results demonstrate that when run on 32 (64) cores, the meshfree method can produce accurate marginalized posteriors for the GW170817 event and locate it in the sky within ~ 3.4 (2.4) minutes of detecting the event. The resulting posteriors obtained from the meshfree method are statistically indistinguishable from those obtained using the brute-force method implemented in PyCBC [180]. Please note that the brute force method is not optimized and is used only as a yardstick for accuracy.

The rest of the paper is structured into the following sections: Section 4.2 introduces the PE basics, where we briefly discuss the Bayesian approach for inferring GW parameters from the data and define the coherent network likelihood. Section 4.3 explains the start-up and online stages and discusses the likelihood evaluation procedure using meshfree interpolation. The next section 4.4.1 presents the results of a detailed analysis of the meshfree method for the GW170817 BNS event. In Section 4.4.2, we examine the effectiveness of our method by testing it on simulated events covering a wide range of signal-to-noise ratios (SNR). Finally, we summarize the results in Section 4.5 and discuss the limitations and future follow-ups of the current implementation. Finally, we summarize the results in Section 4.5 and discuss the limitations of the current implementation (such as restricting the sampler over narrow prior bounds). We suggest some ideas for overcoming this limitation in follow-up studies.

4.2 Parameter estimation

4.2.1 Bayesian inference

Given a stretch of GW strain data \mathbf{d} from detectors containing a GW signal $\mathbf{h}(\vec{\Lambda})$, embedded in additive Gaussian noise \mathbf{n} , we want to estimate the posterior distribution $p(\vec{\Lambda} | \mathbf{d})$ over the

source parameters $\vec{\Lambda}$. The posterior distribution, in turn, is related to the likelihood function $\mathcal{L}(\mathbf{d} | \vec{\Lambda})$ via the well-known Bayes' theorem:

$$p(\vec{\Lambda} | \mathbf{d}) = \frac{\mathcal{L}(\mathbf{d} | \vec{\Lambda}) p(\vec{\Lambda})}{p(\mathbf{d})} \quad (4.1)$$

where, $p(\vec{\Lambda})$ is the prior distribution over model parameters $\vec{\Lambda} \equiv \{\vec{\lambda}, \vec{\theta}, t_c\}$. Here $\vec{\lambda}$ is a set of intrinsic parameters such as the component masses $m_{1,2}$ and dimensionless aligned spins $\chi_{1z,2z}$ while $\vec{\theta}$ represents the set of extrinsic parameters such as the source's sky location i.e. right ascension (α) and declination (δ), the inclination (ι) of the orbital plane of the binary with respect to the line of sight, the polarization angle (ψ), the luminosity distance (d_L) of the source from the Earth and the geocentric epoch of coalescence t_c .

In principle, given a numerical prescription for calculating the likelihood function $\mathcal{L}(\mathbf{d} | \vec{\Lambda})$ under a waveform model and assume prior distributions over the model parameters, we can evaluate the left-hand side of Eq. (4.1) at any given point in the sample space upto an overall normalization factor. However, with a large number of parameters, $\vec{\Lambda}$ (typically ~ 15 parameters), evaluating $p(\vec{\Lambda} | \mathbf{d})$ on a fine grid over the sample space becomes increasingly tedious and eventually intractable with finite computational resources. Therefore, a more intelligent strategy is employed where one uses stochastic sampling techniques to estimate the posterior distribution $p(\vec{\Lambda} | \mathbf{d})$. There are a number of schemes to sample the posterior distribution, such as Markov Chain Monte Carlo (MCMC) [143] and its variants, Nested Sampling [71] algorithms, etc. In this paper, we use DYNESTY [72, 73], an extensively used Python implementation of the nested sampling algorithm for GW data analysis. In this work, we have only focused on speeding up the PE algorithm by quickly evaluating the likelihood function at any point proposed by the sampling algorithm. Admittedly, another aspect of designing a fast-PE algorithm would involve optimizing the sampling algorithm itself - which is not considered in this work. It is easy to see that such improvements to the sampling algorithm can positively impact the performance of many fast-PE methods (including ours). For example, a significant improvement in computation time has been demonstrated by efficiently populating the parameter space as proposed by the VARAHA sampling technique [181].

4.2.2 The likelihood function

Let $\mathbf{d}^{(i)}$ be the strain data recorded at the i^{th} detector containing an astrophysical GW signal $\tilde{h}^{(i)}(\vec{\Lambda})$. We pause to remark that complex $\tilde{h}^{(i)}(\vec{\Lambda})$ will denote the frequency-domain Fourier transform (FT) of the signal $h^{(i)}(\vec{\Lambda})$. Assuming uncorrelated noise among the N_d detectors in the network, the coherent log-likelihood [182] is given by

$$\ln \mathcal{L}(\vec{\Lambda}) = \sum_{i=1}^{N_d} \langle \mathbf{d}^{(i)} | \tilde{h}^{(i)}(\vec{\Lambda}) \rangle - \frac{1}{2} \sum_{i=1}^{N_d} \left[\|\tilde{h}^{(i)}(\vec{\Lambda})\|^2 + \|\mathbf{d}^{(i)}\|^2 \right]. \quad (4.2)$$

For a non-precessing GW signal model, one can use the relation $\tilde{h}_x \propto -j \tilde{h}_+$ between the two polarization states to express the signal at the i^{th} detector as:

$$\tilde{h}^{(i)}(\vec{\Lambda}) \equiv \tilde{h}(\vec{\Lambda}, t^{(i)}) = \mathcal{A}^{(i)} \tilde{h}_+(\vec{\lambda}, t^{(i)}), \mathcal{A}^{(i)} \tilde{h}_+(\vec{\lambda}) e^{-j 2\pi f t_c} e^{-j 2\pi f \Delta t^{(i)}}. \quad (4.3)$$

The complex amplitude of the signal $\mathcal{A}^{(i)}$ depends only on the extrinsic parameters $\vec{\theta} \in \vec{\Lambda}$ through the antenna pattern functions, luminosity distance, and the inclination angle:

$$\mathcal{A}^{(i)} = \frac{1}{d_L} \left[\frac{1 + \cos^2 \iota}{2} F_+^{(i)}(\alpha, \delta, \psi) - j \cos \iota F_\times^{(i)}(\alpha, \delta, \psi) \right] \quad (4.4)$$

where $F_+^{(i)}(\alpha, \delta, \psi)$ and $F_\times^{(i)}(\alpha, \delta, \psi)$ are respectively the ‘plus’ and ‘cross’ antenna pattern functions of the i^{th} detector. The antenna pattern functions of a detector describe the angular response of the detector to incoming GW signals. It arises from contracting the position and geometry dependent ‘detector tensor’ [183] with the metric perturbations (GW); thereby mapping the latter to a GW strain amplitude $h^{(i)}(\vec{\Lambda})$ recorded at the detector.

As indicated by the Eq. (4.3), the signal acquires an additional phase difference during its projection from the geocentric frame to the detector frame, which corresponds to a time delay denoted by $\Delta t^{(i)}$. This temporal offset originates due to the relative positioning of the i^{th} detector in relation to the Earth’s center, and it can be expressed explicitly as follows:

$$\Delta t^{(i)} \equiv t^{(i)} - t_c = \frac{\vec{x}^{(i)} \cdot \hat{N}(\alpha, \delta)}{c}, \quad (4.5)$$

where $\vec{x}^{(i)}$ is a vector pointing from the the Earth’s centre to the location of the i^{th} detector, $t^{(i)}$ is the time at the i^{th} detector, and $\hat{N}(\alpha, \delta)$ is the direction of the GW propagation [184]. In our analysis, we consider the log-likelihood function marginalized over the coalescence phase [145] parameter, which can be written as follows:

$$\ln \mathcal{L}(\vec{\Lambda} \mid \mathbf{d}^{(i)}) \Big|_{\phi_c} = \ln I_0 \left[\left| \sum_{i=1}^{N_d} \langle \mathbf{d}^{(i)} \mid \tilde{h}^{(i)}(\vec{\Lambda}) \rangle \right| \right] - \frac{1}{2} \sum_{i=1}^{N_d} \left[\|\tilde{h}^{(i)}(\vec{\Lambda})\|^2 + \|\mathbf{d}^{(i)}\|^2 \right], \quad (4.6)$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind. By marginalizing over extrinsic parameters, we effectively reduce the dimensionality of the problem, resulting in accelerated likelihood calculations and enhanced sampling convergence. Substituting Eq. (4.3) in the above equation, we get

$$\begin{aligned} \ln \mathcal{L}(\vec{\Lambda} \mid \mathbf{d}^{(i)}) \Big|_{\phi_c} &= \ln I_0 \left[\left| \sum_{i=1}^{N_d} \mathcal{A}^{(i)*} \langle \mathbf{d}^{(i)} \mid \tilde{h}_+(\vec{\lambda}, t^{(i)}) \rangle \right| \right] \\ &\quad - \frac{1}{2} \sum_{i=1}^{N_d} \left[\left| \mathcal{A}^{(i)} \right|^2 \sigma^2(\vec{\lambda})^{(i)} + \|\mathbf{d}^{(i)}\|^2 \right]; \end{aligned} \quad (4.7)$$

where, $\langle \mathbf{d}^{(i)} \mid \tilde{h}_+(\vec{\lambda}, t^{(i)}) \rangle$ is the complex overlap integral, while $\sigma^2(\vec{\lambda})^{(i)} \equiv \langle \tilde{h}_+(\vec{\lambda}, t^{(i)}) \mid \tilde{h}_+(\vec{\lambda}, t^{(i)}) \rangle$ is the squared norm of the template $\tilde{h}_+(\vec{\lambda})$. $\sigma^2(\vec{\lambda})^{(i)}$ depends on the noise power spectral density (PSD) of the i^{th} detector. The squared norm of the data vector, $\|\mathbf{d}^{(i)}\|^2$, remains constant during the PE analysis and, therefore, doesn’t impact the overall ‘shape’ of the likelihood. As such, it can be excluded from the subsequent analysis.

4.3 Meshfree likelihood interpolation

As discussed in our previous work [179], the meshfree method comprises two distinct stages:

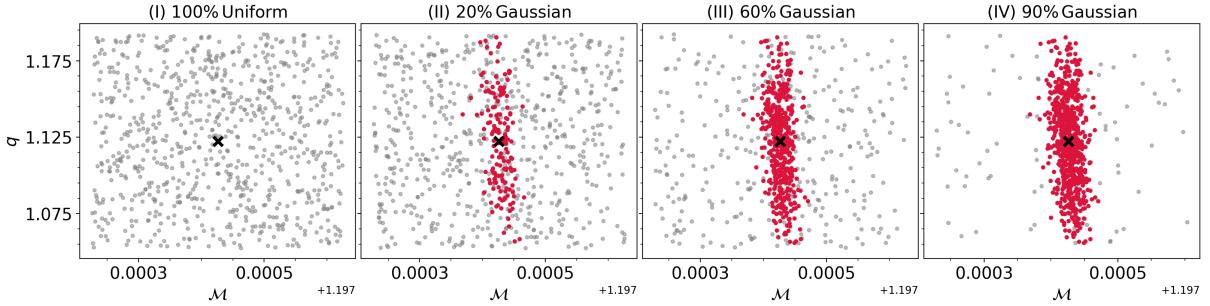


Figure 4.1: The figure displays randomly selected RBF interpolation nodes in the intrinsic parameter space with varying fractions from Gaussian and uniform distributions (2D slice in \mathcal{M} and q coordinates). Randomly placed nodes drawn from a Gaussian distribution are denoted in red, while those from a uniform distribution are shown in gray. The covariance matrix of the Gaussian distribution can be estimated semi-analytically given the signal model. Sub-figure-(I) represents a choice with 100% (all) nodes drawn from a uniform distribution, and the other sub-figures exhibit mixtures of different percentages of nodes from the uniform and Gaussian distributions. The center (marked as a “black cross”) of the sample space is determined through an optimization routine based on the highest network signal-to-noise ratio (SNR). This search is guided by the best-matched trigger obtained from the upstream search pipeline.

1. Start-up Stage: During this phase, we create radial basis function (RBF) interpolants for the pertinent quantities.
2. Online Stage: In this stage, we swiftly compute the likelihood function at query points proposed by the sampling algorithm.

Next, we will delve into the various stages that constitute meshfree interpolation.

4.3.1 Start-up stage

During the start-up stage, we generate RBF interpolants for the relevant quantities, which enable rapid likelihood calculations in the online stage. The start-up stage can be further divided into the following parts:

- i. Placement of RBF nodes over the intrinsic parameter space, λ . These discrete nodes are denoted as $\vec{\lambda}^n : n = 1, 2, \dots, N$.
- ii. Generation of complex time-series $\vec{z}^{(i)}(\vec{\lambda}^n) \equiv \langle \mathbf{d}^{(i)} | \tilde{h}_+(\vec{\lambda}^n, t^{(i)}) \rangle$ and template norm square $\sigma^2(\vec{\lambda}^n)^{(i)}$ at each of the RBF nodes $\vec{\lambda}^n$, for each detector. In the start-up stage, we calculate the $\vec{z}^{(i)}(\vec{\lambda}^n)$ at discrete values of circular time shifts t_c (with respect to some reference geocentric coalescence time) uniformly spaced over a specified range. The range of time shifts $|t_c|$ should be larger than the maximum light travel time between two detectors (corresponding to the longest baseline). For the analysis presented here we choose the time shifts in the range ± 150 ms around the geocentric time of coalescence as reported by an upstream search pipeline. Note that we set $\Delta t^{(i)} = 0$ for all the detectors while calculating $\vec{z}^{(i)}(\vec{\lambda}^n)$. Additional time

delays $\Delta t^{(i)}$ for the signal to reach the i^{th} detector depending on the source's location in the sky can be taken care of during the online stage while sampling the posterior distribution over the sky location parameters as explained later.

- iii. Singular value decomposition (SVD) of the complex time-series matrix $\mathcal{Z}^{(i)}$ at the i^{th} detector, where the n^{th} row is defined as $\vec{z}^{(i)}(\vec{\lambda}^n)$, and
- iv. Generation of RBF interpolating functions for the SVD coefficients and the template norm square σ^2 for each detector.

We now describe some of these steps in greater detail.

RBF nodes placement

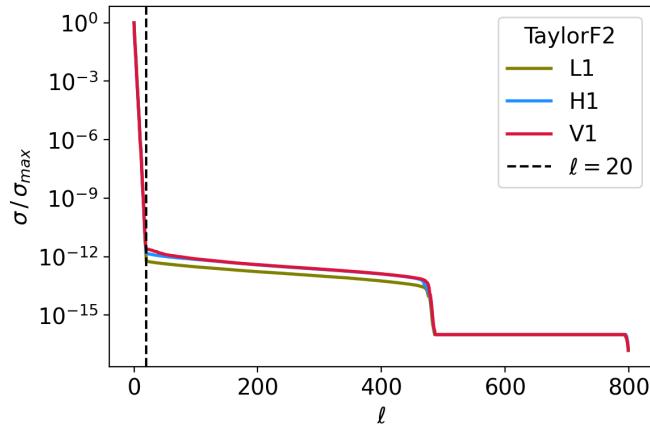


Figure 4.2: The figure shows the profile of the singular values normalized by the highest singular value. Notably, there is a sharp decline in the singular values (indicated by a black dashed line) due to the large correlation between the time-series $\vec{z}^{(i)}(\vec{\lambda}^n)$ at different RBF interpolation nodes, revealing that only the linear combination of the top-few singular vectors (around 20) is sufficient for effectively reconstructing the time series $\vec{z}^{(i)}(\vec{\lambda}^q)$ with minimal reconstruction error. The 3.5 PN TaylorF2 post-Newtonian signal model is assumed.

The accurate construction of meshfree RBF interpolants relies significantly on the strategy of randomly distributing the RBF nodes, denoted as $\vec{\lambda}^n$, across the sample space. However, the task of finding an optimal set of nodes poses a considerable challenge. Typically, a hypercube in the sample space is selected around a reference point during the sampling of the posterior distribution. This reference point commonly known as trigger or best-matched template $\vec{\lambda}^*$ and trigger time (geocentric) t_{trig} are derived from an upstream search pipeline [141, 15].

In our previous work, we opted for a uniform distribution of nodes within the hypercube centered around the reference point (injection in the case of a simulated BNS event). This approach works well for events with low SNRs since the corresponding likelihood function exhibited a relatively flatter profile across the parameter space. However, for high SNR events, where the likelihood profile exhibits a sharp peak in certain regions of the parameter space, uniformly distributing nodes within the hypercube may not adequately cover the support of

the posterior distribution, particularly in areas where we seek higher accuracy in interpolated likelihood estimates. As a result, we adopted a combination of nodes drawn from both a multivariate Gaussian distribution and a uniform distribution across the sample space within the hypercube. The multivariate Gaussian distribution is defined by a mean vector $\vec{\mu}$ and a covariance matrix Σ , and its expression is as follows:

$$p(\vec{\lambda}) \propto \exp \left[-(\vec{\lambda} - \vec{\mu})^T \Sigma^{-1} (\vec{\lambda} - \vec{\mu}) \right] \quad (4.8)$$

where μ can be selected as the trigger point $\vec{\lambda}^*$, and Σ can be computed using the inverse of the Fisher matrix Γ evaluated at $\vec{\lambda}^*$. While the distribution of nodes generated using the Fisher matrix may not perfectly match the actual posterior distribution across the sample space, it does exhibit significant overlap with the true distribution, especially in the vicinity of the likelihood peak.

Since the trigger originates from a search pipeline that uses a discrete set of templates over a grid (also known as a template bank) [185, 186], we can select a reference point, denoted as $\vec{\lambda}_{\text{ref}}$, in close proximity to the trigger point ($\vec{\lambda}^*$), which may have a higher network SNR. We employ an optimization strategy [187] aimed at maximizing the network SNR in order to identify this reference point. Subsequently, a portion of the interpolation nodes are generated from a multivariate Gaussian distribution, $\mathcal{N}(\vec{\lambda}_{\text{ref}}, \Sigma)$, where Σ is computed from the inverse of the Fisher matrix evaluated at $\vec{\lambda}_{\text{ref}}$, to adequately cover the region around the peak. For regions of the sample space where Gaussian nodes are sparse or possibly nonexistent, we spray a uniform distribution of samples across the hypercube. This approach ensures high accuracy in interpolated likelihood estimates throughout the entire sample space.

For a visual representation illustrating the various combinations of Gaussian and uniform nodes, please refer to Figure 4.1. In particular, note that choosing a small fraction of random RBF nodes from the multivariate Gaussian distribution leads to more accurate posterior distributions over certain parameters (see Appendix 4.6.1).

SVD of $\mathcal{Z}^{(i)}$ and Interpolants generation

After obtaining the appropriate RBF nodes $\vec{\lambda}^n$, we can efficiently compute the time-series $\vec{z}^{(i)}(\vec{\lambda}^n) \equiv z^{(i)}(\vec{\lambda}^n, t_c)$ using Fast Fourier Transform (FFT) based circular correlations where t_c are the uniformly spaced discrete-time shifts taken from a specified range around a reference coalescence time t_{trig} . Given that the predominant portion of the posterior support originates from the vicinity of the likelihood peak, we opt for $t_c \in [t_{\text{trig}} \pm 0.15 \text{ s}]$ as the interval for sampling t_c .

Similarly, we compute the template norm square $\sigma^2(\vec{\lambda}^n)^{(i)}$ at all the RBF nodes $\vec{\lambda}^n$.

Our objective is to identify the basis vectors capable of spanning the space of $\vec{z}^{(i)}(\vec{\lambda}^n)$. This can be achieved by stacking (row-wise) the time-series vectors $\vec{z}^{(i)}(\vec{\lambda}^n)$ for each $\vec{\lambda}^n$, where $n = 1, 2, \dots, N$. Subsequently, performing the Singular Value Decomposition (SVD) on the resultant matrix $\mathcal{Z}^{(i)}$ yields our desired basis vectors \vec{u}_μ . It can be shown that any vector $\vec{z}^{(i)}(\vec{\lambda}^n)$ within the specified parameter range can be expressed as a linear combination of these basis

vectors \vec{u}_μ ,

$$\vec{z}^{(i)}(\vec{\lambda}^n) = \sum_{\mu=1}^N C_\mu^{n(i)} \vec{u}_\mu^{(i)} \quad (4.9)$$

where, $C_\mu^{n(i)} \equiv C_\mu^{(i)}(\vec{\lambda}^n)$, $\mu = 1, 2, \dots, N$ are the N SVD coefficients. These coefficients $C_\mu^{n(i)}$ are characterized as smooth functions (see Appendix 4.6.3) over $\vec{\lambda}^n$ within a sufficiently narrow boundary encompassing the posterior support. Therefore, each of these N coefficients can be treated as a scalar-valued, smooth function over the d -dimensional intrinsic parameter space and expressed as a linear combination of the radial basis functions [107] (RBFs) ϕ centered at the interpolation nodes:

$$C_\mu^{q(i)} = \sum_{n=1}^N a_n^{(i)} \phi(\|\vec{\lambda}^q - \vec{\lambda}^n\|_2) + \sum_{j=1}^M b_j^{(i)} p_j(\vec{\lambda}^q) \quad (4.10)$$

where ϕ is the RBF kernel centered at $\vec{\lambda}^n \in \mathbb{R}^d$, and $\{p_j\}$ denotes the monomials that span the space of polynomials with a predetermined degree ν in d -dimensions. The inclusion of these monomial terms has been shown to enhance the accuracy of RBF interpolation [108].

Similarly, we can also express $\sigma^2(\vec{\lambda}^q)^{(i)}$ in terms of the RBF functions and monomials by treating it as smoothly varying scalar-valued function over the interpolation domain.

From Eq. (4.10), we find that there are a total of $(N + M)$ coefficients to be solved for each interpolating function. The SVD coefficients $C_\mu^{q(i)}$ and $\sigma^2(\vec{\lambda}^q)^{(i)}$ are known at each of the N RBF nodes $\vec{\lambda}^n$. These provide N interpolation conditions. To uniquely determine the coefficients a_n and b_j , we impose M additional conditions of the form $\sum_{j=1}^M a_j^{(i)} p_j(\vec{\lambda}^q) = 0$. This leads to the following system of $N + M$ equations:

$$\begin{bmatrix} \Phi & P \\ P^T & O \end{bmatrix} \begin{bmatrix} a^{(i)} \\ b^{(i)} \end{bmatrix} = \begin{bmatrix} C_\mu^{n(i)} \\ 0 \end{bmatrix} \quad (4.11)$$

where the matrices Φ and P have components $\Phi_{ij} = \phi(\|\vec{\lambda}^i - \vec{\lambda}^j\|_2)$ and $P_{ij} = p_j(\vec{\lambda}^i)$ respectively; $O_{M \times M}$ is a zero-matrix and $0_{M \times 1}$ is a zero-vector.

The additional M conditions arise due to the fact that Φ needs to be a real, symmetric, and conditionally positive definite matrix of order $\nu + 1$, and the following theorem (reproduced here from [107] for completeness) guarantees the uniqueness of the solutions:

Theorem 1 *Let A be a real symmetric $N \times N$ matrix that is conditionally positive definite of order one, and let $B = [1, \dots, 1]^T$ be an $N \times 1$ matrix (column vector). Then the system of linear equations*

$$\begin{bmatrix} A & B \\ B^T & O \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} y \\ 0 \end{bmatrix}$$

is uniquely solvable.

Although Theorem 1 is applicable for cases where $\nu = 0$, it can be generalized to polynomials of higher degrees (For a proof, refer to Theorem 7.2 in [107]).

Now, the set of equations Eq. (4.11) can be uniquely solved for the unknown coefficients $a^{(i)}$ and $b^{(i)}$ to determine the meshfree interpolating functions corresponding to the coefficients $C_\mu^{(i)}$.

From the sharply decreasing spectrum of eigenvalues shown in Fig. 4.2, it is clear that only the “top-few” basis vectors are sufficient to reconstruct $\vec{z}^{(i)}(\vec{\lambda}^q)$ at minimal reconstruction error, we generate only top- ℓ meshfree interpolants of $C_\mu^{q(i)}$ where $\mu = 1, \dots, \ell$. Note that ℓ is chosen based on the spectrum of eigenvalues of the matrix $\mathcal{Z}^{(i)}$. We have chosen $\ell = 20$ where the normalized eigenvalues fall (see Fig. 4.2) to $\sim 10^{-12}$. However, this is a subjective choice.

In a similar way, we generate the meshfree interpolants for the scalar-valued function $\sigma^2(\vec{\lambda}^q)^{(i)}$ as well.

Note that to optimize computational efficiency, we need to generate a minimum number of RBF nodes $\vec{\lambda}^n$ given the degree (ν) of the polynomial and dimension (d) of the intrinsic parameter space and it is given by $N_{\min} = \binom{\nu+d}{\nu}$. Increasing the number of nodes increases not only the computational cost of generating the interpolants but also their evaluation cost during the online stage, which is evident from Eq. (4.10) as the number of coefficients $a_n^{(i)}$ increase linearly with the number of RBF nodes N . In our study, we empirically choose at least twice (or more) the minimum number of required nodes.

At the end of the start-up stage, we thus have uniquely determined $(\ell + 1)$ interpolating functions

4.3.2 Online stage

Once we have the meshfree interpolants ready, we can evaluate the meshfree interpolated values of $C_\mu^{q(i)}$ and $\sigma^2(\vec{\lambda}^q)^{(i)}$ at any query point $\vec{\lambda}^q$ rapidly. We then combine the interpolated values of the coefficients $C_\mu^{q(i)}$ with the corresponding basis vectors \vec{u}_μ according to Eq. (4.9), restricting the summation to the first ℓ terms.

In the startup stage 4.3.1, we set $\Delta t^{(i)} = 0$ while calculating $\vec{z}^{(i)}(\vec{\lambda}^n)$ and didn’t include the time-delay relative to the Earth’s center, which depends on the sky location. Now, we incorporate it while reconstructing the interpolated time-series $\vec{z}^{(i)}(\vec{\lambda}^q)$. Moreover, we focus on constructing only $\vec{z}^{(i)}(\vec{\lambda}^q)$ at approximately 10 time-samples centered around query time $t^{q(i)}$ (Eq. (4.5)). To ensure sub-sample accuracy, we fit these samples with a cubic spline, which incurs negligible computational cost. Subsequently, we calculate $\vec{z}^{(i)}(\vec{\lambda}^q)$ at the query $t^{q(i)}$ using the cubic-spline interpolant. Similarly, we evaluate the interpolated value of $\sigma^2(\vec{\lambda}^q)^{(i)}$. Finally, combining $\vec{z}^{(i)}(\vec{\lambda}^q)$ and $\sigma^2(\vec{\lambda}^q)^{(i)}$ with coefficients dependent on extrinsic parameters $\vec{\theta}$, as exemplified in Eq. (4.7), enables the computation of $\ln \mathcal{L}_{\text{RBF}}$.

4.4 Numerical experiments

4.4.1 GW170817 event

We conducted a Bayesian PE study on the GW170817 binary neutron star (BNS) event to test our method. The strain data from the two advanced-LIGO and Virgo detectors were obtained from the open archival data sets available on GWOSC [188, 189]. We used a 360 seconds data segment that was cleaned with a 4th-order high-pass filter, with a cut-off frequency of 18 Hz. The seismic cutoff frequency was set to 20 Hz. The TaylorF2 waveform model [98, 99, 100, 101] was employed to recover the source parameters. The noise PSD was estimated from the strain

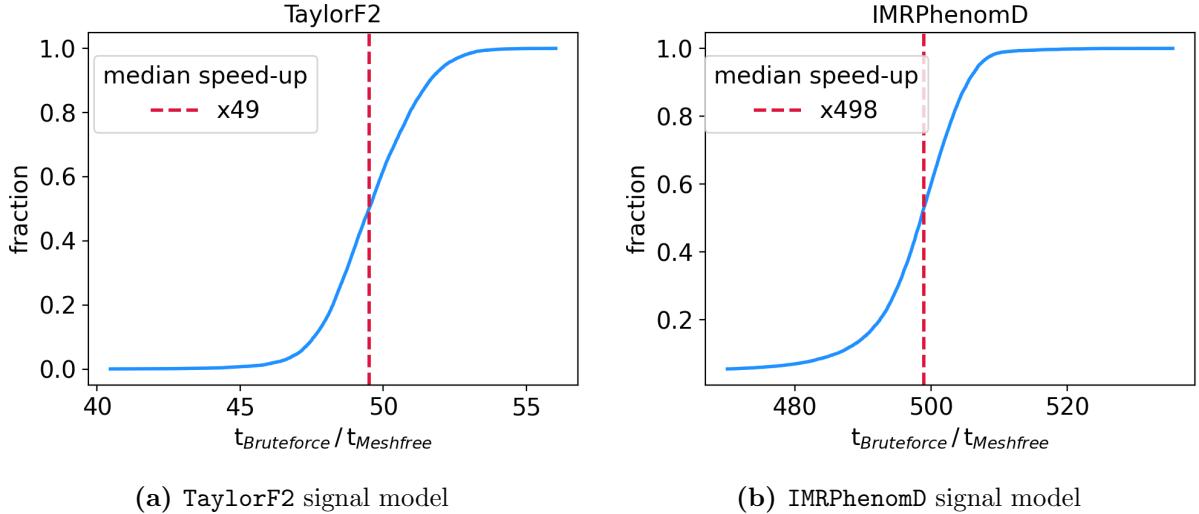


Figure 4.3: The plots compare the speed-up of the meshfree likelihood evaluation with bruteforce likelihood for two waveform models, `TaylorF2` and `IMRPhenomD`. The meshfree method is around 49 times faster for `TaylorF2` and approximately 498 times faster for `IMRPhenomD` (compared to the standard calculation). The median absolute error is approximately $\mathcal{O}(10^{-2})$ for both cases. In the proposed meshfree method, one can trade-off speed in favor of better accuracy by retaining a larger number of SVD basis vectors (see Eq. (4.9)) and vice-versa. The optimal subjective choice is made from the spectrum of eigenvalues (Fig. 4.2).

data using overlapping segments of length 2 seconds, using the `median-mean` PSD estimation method by applying a `Hann` window as implemented in `PYCBC`. Subsequently, we generated $N = 800$ RBF nodes using the node placement algorithm described in section 4.3.1.

Out of total RBF nodes, $N_{\text{Gauss}} = 200$ nodes (number of Gaussian nodes) are selected from a multivariate Gaussian distribution, $\mathcal{N}(\vec{\lambda}_{\text{ref}}, \Sigma)$ where the mean is set as $\vec{\lambda}_{\text{ref}}$, and the covariance matrix, Σ is estimated by evaluating inverse of the Fisher matrix, Γ [149] at the reference point $\vec{\lambda}_{\text{ref}}$ (see Section 4.3.1). We utilize the `gwfast`[190] Python package to calculate the covariance matrix. The remaining nodes $N_{\text{Unif}} = 600$ are uniformly sampled from the ranges specified in Table 4.1. For generating the RBF interpolants, we employ the publicly available `RBF` python package [118]. A Gaussian RBF kernel $\phi = \exp(-\varepsilon r^2)$ is used as the basis function, where ε serves as the shape parameter. We set $\varepsilon = 10$ through trial and error, but it can be optimized using the leave-one-out cross-validation [191] method. The degree of the monomials is chosen as $\nu = 7$, and we retain the top 20 basis vectors ($\ell = 20$) for the reconstruction of $\vec{z}^{(i)}(\vec{\lambda}^q)$.

To sample the ten-dimensional parameter space $\vec{\Lambda}$, we utilize nested sampling implemented in the `dynesty` python package. The likelihood function used for the analysis is $\ln \mathcal{L}_{\text{RBF}}$. The prior distributions and their boundaries for all sampling parameters are provided in Table 4.1. The sampler configuration is outlined as follows: `nlive` = 500, `walks` = 100, `sample` = “rwarm”, and `dlogz` = 0.1. These parameters crucially determine the accuracy and time taken by the nested sampling algorithm. Here, `nlive` represents the number of live points. Opting for a larger number of `nlive` points results in a more finely sampled posterior (and consequently,

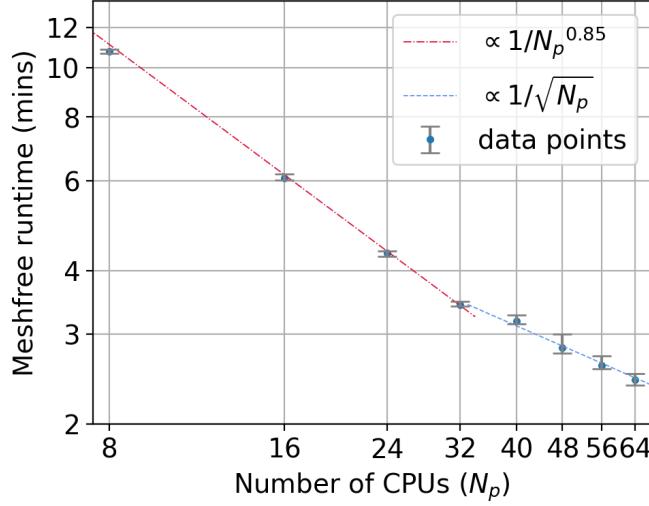


Figure 4.4: The runtimes of the meshfree method have been plotted with varying the number of cores. We choose the GW170817 event for this specific example, and TaylorF2 waveform model is taken for obtaining the posteriors. The plot shows a notable decrease in meshfree runtimes as the number of cores increased. We anticipate significant reductions in runtimes can be achieved with an even higher number of cores, though the gains will eventually saturate as we increase the number of cores.

evidence), albeit at the expense of requiring more iterations for convergence. `walks` denotes the minimum number of points needed before proposing a new live point, `sample` indicates the chosen sampling approach for generating samples, and `dlogz` signifies the proportion of the remaining prior volume’s contribution to the total evidence, which functions as a stopping criterion for terminating the sampling process. Further details on the dynesty’s nested sampling algorithm and its implementation can be found in [72, 73]. For comparison, we also employ the bruteforce marginalized phase likelihood implemented in PYCBCwith the dynesty sampler, using the same sampler configuration mentioned above.

The PE run with the meshfree likelihood (for TaylorF2) completed in ~ 3.4 minutes

Parameters	Range	Prior distribution
\mathcal{M}	$[\mathcal{M}_{\text{cent}} \pm 0.0002]$	$\propto \mathcal{M}$
q	$[q_{\text{cent}} \pm 0.07]$	$\propto [(1 + q)/q^3]^{2/5}$
$\chi_{1z,2z}$	$[\chi_{1z \text{ cent}} \pm 0.0025]$	Uniform
d_L	$[10, 60]$	Uniform in volume
t_c	$t_{\text{trig}} \pm 0.12$	Uniform
α	$[0, 2\pi]$	Uniform
δ	$\pm\pi/2$	$\sin^{-1} [\text{Uniform}[-1, 1]]$
ι	$[0, \pi]$	Uniform in $\cos \iota$
ψ	$[0, 2\pi]$	Uniform angle

Table 4.1: Prior parameter space over the ten-dimensional parameter space $\vec{\Lambda}$.

(including the startup and sampling stages) when run on 32 cores. In a similar analysis with the `IMRPhenomD` [102] waveform model, we found good agreement between the posterior distributions obtained from the meshfree and PyCBC likelihood methods, and these results are broadly consistent with the LVK analysis of the same event. In this case, the meshfree method generated the posterior in ~ 3.6 minutes utilizing 32 cores. As shown in Fig. 4.5, the posterior distributions estimated using the meshfree method and bruteforce likelihood exhibit good agreement. Additionally, sky maps for GW170817 were generated using the `ligo-sky-map` [192] utility for two methods, further demonstrating the effectiveness of the meshfree method in reconstructing source parameters. For the same analysis using 64 cores, the number of likelihood evaluations for meshfree PE (standard PyCBC PE) is 1396388 (1404588) whereas the number of posterior samples collected during sampling is 3633 (3546).

To estimate the speed-up achieved by using the meshfree likelihood compared to the bruteforce likelihood, we generated 10^4 random query points from the prior distribution specified in Table 4.1. Subsequently, we calculated the log-likelihood at these points using both the meshfree and bruteforce likelihood methods. The results show a speed-up of ~ 49 over standard bruteforce likelihood, as shown in Fig. 4.3. In the case of `IMRPhenomD`, we observed a relative speed-up of ~ 498 over standard bruteforce likelihood. This difference in speed-up can be attributed to the difference in time taken to generate the template in the two different waveform models: `TaylorF2` waveforms are far less computationally expensive to generate than `IMRPhenomD` model waveforms. The corresponding error plots are shown in Appendix 4.6.2.

Furthermore, we investigated the scaling behavior of PE run times (startup + online) using the meshfree method with respect to the number of cores employed. For the `TaylorF2` analysis of the GW170817 event, we performed the ten-dimensional PE analysis with varying core counts ranging from 8 to 64. Our study indicates that the analysis can be completed in ~ 2.4 minutes using 64 cores. Furthermore, we found that the current implementation scales linearly with the number of live points (keeping `dlogz` fixed).

It is worth noting that further reductions in estimation times are expected if the number of cores is increased, as the scaling does not saturate at 64 cores. However, as depicted in Fig. 4.4, the rate at which the estimation times decrease diminishes as the number of cores continues to increase, which may eventually result in run time saturation beyond a certain number of cores. The total time shown in Fig. 4.4 includes the time taken by both the start-up and online stages of the meshfree algorithm, in addition to the time required for the dynamic nested sampling algorithm to compute the evidence integral up to a desired level of accuracy. To clarify further: (a) The start-up stage of the meshfree method benefits from the availability of many CPU cores as it can be performed in parallel. (b) Conversely, the online stage is not designed for parallel processing. This is because estimating the likelihood is a relatively fast operation, typically taking $\sim \mathcal{O}(1)$ ms, achieved by evaluating the interpolating functions. This part of the computational cost remains nearly constant regardless of the specific sample point being processed. As a result, the runtime scaling, as depicted in Fig. 4.4, effectively combines the time consumed by the Dynesty sampler and the start-up stage of our method.

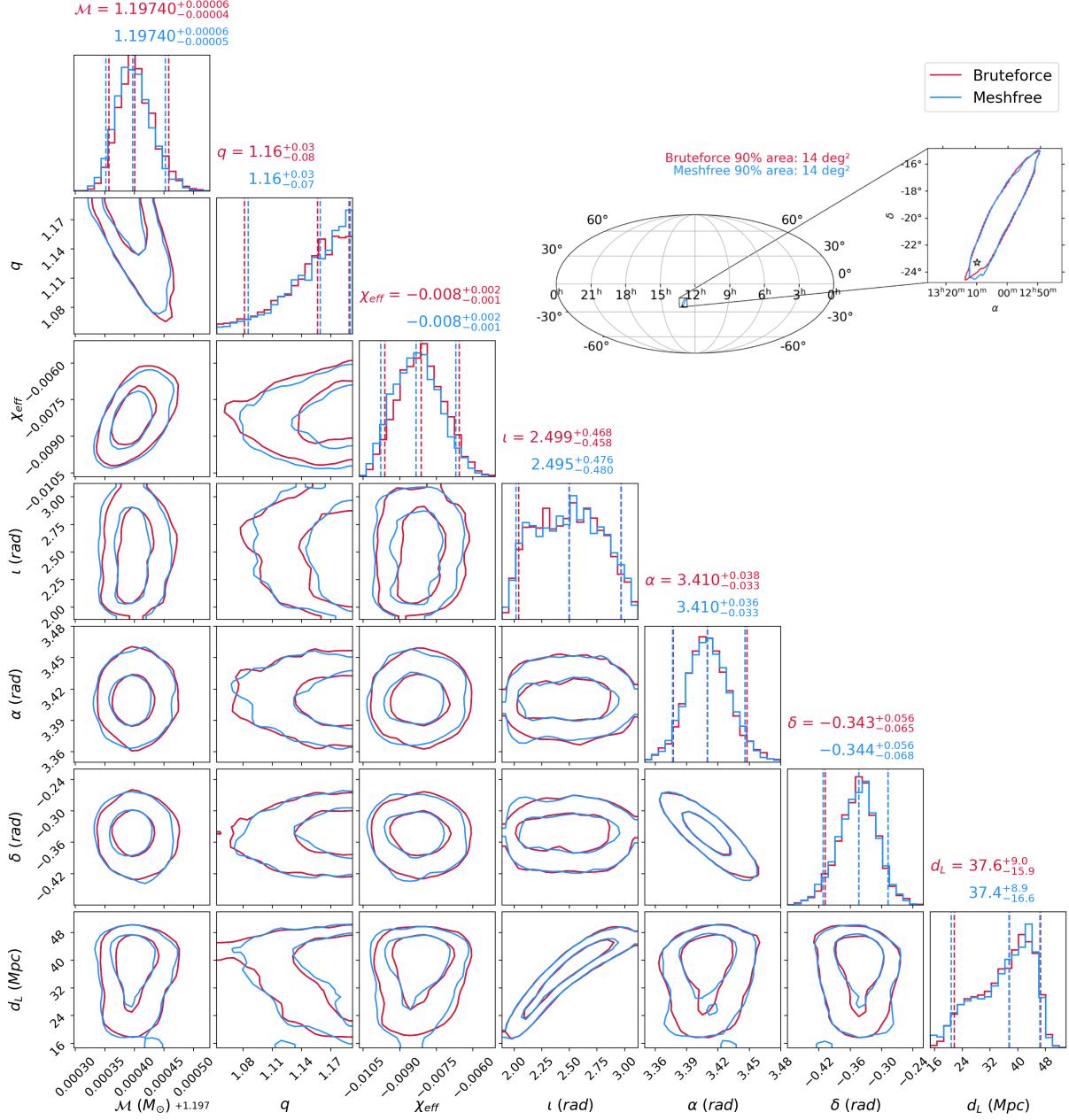


Figure 4.5: The corner plot shows the posterior distributions over a ten-dimensional parameter space using both the meshfree method and the standard brute force likelihood. The seismic cutoff frequency is set to 20 Hz, assuming the TaylorF2 waveform model. The estimated distributions from both methods exhibit good agreement, as indicated by the median values in the title of the 1-D marginalized posterior plots along the diagonal. The meshfree likelihood-based PE process was completed in ~ 3.4 minutes when performed on 32 cores. A skymap (inset) is also generated for both methods where the star denotes the actual location of the galaxy from which GW170817 is believed to have originated and is in agreement with [14].

4.4.2 Simulated Data

To assess the performance of our method across a range of SNRs, we generated simulated data sets with different source parameters that mimic binary neutron star (BNS) systems with network-matched filter SNRs spanning from 10 to 100. In our approach, the number of Gaussian nodes, denoted as N_{Gauss} , for a given source is intricately tied to its SNR. Specifically, the likelihood profile for events with lower SNRs exhibits a relatively broader spread over the parameter space compared to the more sharply peaked likelihood profiles observed for high SNR events. Consequently, for higher SNR events, we select a larger number of nodes from the Gaussian distribution in contrast to lower SNR events. Thus, we opt for different values of N_{Gauss} based on the SNR of each source. We will now detail the distribution of source parameters.

The component masses were uniformly drawn from the interval $[1.2, 2.35]M_{\odot}$, while the magnitude of the dimensionless component spins (aligned with the orbital angular momentum) was uniformly drawn from the interval $[-0.05, 0.05]$. The sources were uniformly distributed in sky location and followed a uniform distribution with d_L ranging from 10 to 200 Mpc. The inclination angle ι of the binary's plane with respect to the line of sight was chosen to be uniformly distributed in $\cos \iota$. The tidal deformability parameter Λ_{tide} of the neutron stars was set to zero for all signals. These signals were simulated using the `TaylorF2` waveform approximant, and the noise PSD was chosen from `aLIGOZeroDetHighPower` [115] for the Livingston and Hanford detectors, and `AdvVirgo` [31] for the Virgo detector, as implemented in `PyCBC`.

It's important to recognize that the distribution of injection parameters we've chosen may not hold any astrophysical relevance. These choices have been made solely for the purpose of testing our method across various SNR scenarios, as our node placement algorithm is intricately linked to the SNRs of these events.

At this point, we employ both the meshfree technique and the relative-binning method (`RELBIN`) for estimating the BNS source parameters. In an ideal scenario, we would select the brute force likelihood method for comparison, but opting for such an approach might result in extended PE run times. Given that `RELBIN` is a well-established and widely used fast PE algorithm within the GW community, recognized for its combination of speed and accuracy, we have opted for this alternative over the standard approach.

For all the simulated events, we chose $N = 800$ RBF nodes. However, N_{Gauss} was empirically chosen depending on the SNR of the events. Heuristically we found that for $\text{SNR} \geq 50$, $N_{\text{Gauss}} \in [0.4, 0.6]$ was a suitable choice. While for events with $\text{SNR} \leq 50$, N_{Gauss} was chosen from $[0, 0.2]$. There were few exceptions to this prescription as well.

For the `RELBIN` method, we set $\epsilon = 0.000198$, which determines the number of frequency bins in the `RELBIN` method (For more details, refer to [139, 138]). This parameter determines the accuracy and run-time of the relative binning method. We choose the reference parameters (or “fiducial point” in the case of `RELBIN`) to be the injection parameters of each simulated event, even though in a realistic scenario, it would have been obtained by an optimization process starting from the best-matched search template.

The prior distributions over the extrinsic parameters $(\alpha, \delta, \iota, \psi, d_L)$ used in the PE are the same as those used for the GW170817 event in the earlier section. As for the intrinsic parameters,

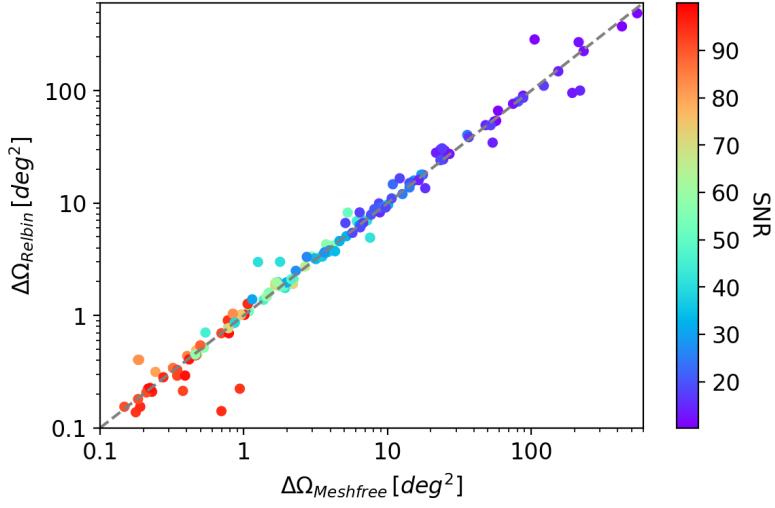


Figure 4.6: 90% credible interval (CI) sky area as a function of SNR are shown. We compared the accuracy of meshfree with the RELBIN method. The scattered points lie along the positive slope (dashed straight line), implying that the meshfree method has similar accuracy as the RELBIN in estimating these parameters, i.e., sky location. It is observed that estimating these parameters with high SNR events is more accurate than with low SNR events. Note that there are some points that are away from the straight line. This is due to the fact that posteriors corresponding to those points are multimodal, giving rise to higher areas for one method in comparison to the other method.

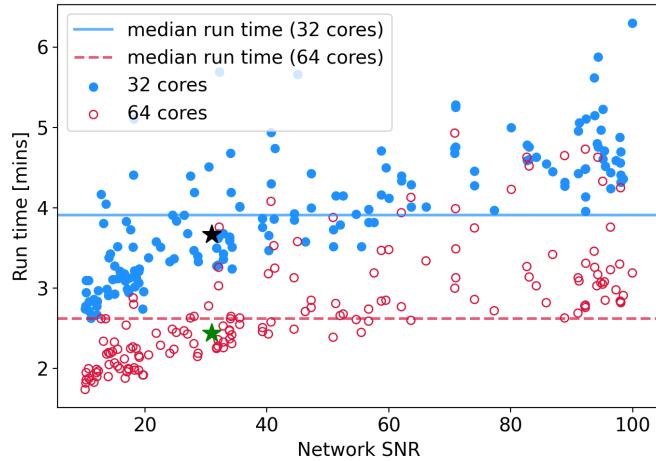


Figure 4.7: We present the run times for reconstructing a number of simulated BNS events injected in Gaussian noise (Sec. 4.4.2) using the meshfree method. Higher SNR injections took longer to reconstruct due to their narrowly peaked likelihood. The analysis utilized 32(64) cores, represented by blue dots (red circles). The black(green) star shows the GW170817 event's run time on 32(64) cores for reference. With 64 cores, the median run time decreased to around ~ 2.6 min compared to ~ 3.9 min with 32 cores. For RELBIN method, using $\epsilon = 0.000198$, the median run time is ~ 22 minutes on 32 cores. The RELBIN PE run times crucially depend on the value ϵ , and taking higher values of ϵ can result in lower run times.

we select them according to the following distributions: (i) Uniform distribution in component masses (m_1, m_2) such that $\mathcal{M} \in \mathcal{M}_{\text{ref}} \pm 0.0002$ and $q \in q_{\text{ref}} \pm 0.07^1$ (ii) Uniformly in dimensionless component spins from the interval $\chi_{1z/2z} \in \chi_{1z/2z_{\text{ref}}} \pm 0.0025$ where $\mathcal{M}_{\text{ref}}, q_{\text{ref}}, \chi_{1z/2z_{\text{ref}}}$ are chirp mass, mass-ratio, and aligned spins components respectively, corresponding to the reference point found by the optimizing the network-SNR starting from the best-matched template. The prior over the time of coalescence t_c is chosen to be uniform from the interval $t_{\text{ref}} \pm 0.12$ where t_{ref} is the geocentric time of the trigger.

As can be seen from the 90% CI sky-area distribution in Fig. 4.6, the meshfree method accurately recovers the sky location with a similar level of accuracy as RELBIN. The timing results for the simulated signals, shown in Fig. 4.7, illustrate that the run-time increases with the SNR since signals with higher SNRs exhibit a more sharply peaked likelihood function that requires a longer time to accumulate a sufficient number of posterior samples to evaluate the evidence integral upto the desired accuracy.

All the tests were performed on AMD EPYC 7542 CPU@2.90GHz processors.

4.5 Conclusion and Outlook

This work extends our previously prescribed meshfree method [179] to perform parameter estimation on a ten-dimensional parameter space for aligned spin waveform models using data from a multi-detector network framework. In the previous version of our method, we demonstrated its effectiveness on a simulated BNS event, as seen in a single detector, where we directly chose the injection parameters as our center for placing uniformly distributed RBF nodes. In this work, we start from the best-matched template parameters and optimize the network-SNR to reach a nearby point with a higher network-SNR value, which acts as a guide for placing RBF nodes. This node placement algorithm incorporates a blend of nodes from both multivariate Gaussian and uniform distributions. It enhances the accuracy of likelihood reconstruction (see Appendix 4.6.1 for a study of marginalized posterior profiles on different fractions of random interpolation nodes drawn from a multivariate Gaussian distribution).

We tested our method for several simulated BNS events with varying SNRs and found good agreement between the meshfree and the relative binning methods. Further, we have demonstrated that our meshfree method can rapidly estimate the posteriors of the GW170817 binary neutron star (BNS) event and efficiently locate the electromagnetic (EM) counterpart in the sky within ~ 2.4 (3.4) minutes after detection using 64 (32) CPU cores.

These run times quoted above need to be put in the context of other fast-PE algorithms for which we chose the relative-binning method. Note that the time taken by the relative binning algorithm crucially depends on the choice of the ϵ parameter, which sets the frequency-bin resolution in the heterodyning process. We set $\epsilon = 0.000198$ for which the time taken by the relative-binning method was found to take ~ 20 minutes on 32 cores to analyze the GW170817 event using data from a 3-detector (HLV) network. This choice of ϵ was taken to match the

¹In cases where $q_{\text{ref}} - 0.07 < 1$, we sample from $q \in [1, 1.14]$.

PE run times claimed by Finstad and Brown (2020) [129]. The sampler parameters were kept identical for both methods. Since we do not have the hardware (nor know the parameters) used by Finstad and Brown, this ad-hoc procedure was used to arrive at this reference time. As such, it won't be prudent to arrive at any definite comparative conclusion based on these timing results. Nevertheless, we will continue to tune the meshfree method to explore further optimizations for speeding up the PE analysis.

Note that the run time quoted here doesn't include the time spent in finding the center using optimization, which, when run on a single core, took ~ 32 seconds. However, it can be sped up by running in parallel over multiple CPU cores using multiprocessor versions of the optimization algorithm. For example, see the parallel version of the `scipy.optimize` [193, 194] Python routine.

We also used our method for the NSBH event GW200115 [195] (See Appendix 4.6.5) and obtained a reasonably good match of posterior samples between our method and bruteforce likelihood. Our meshfree framework has the potential to contribute to future LIGO low-latency PE efforts. However, it is also essential to acknowledge some limitations of the current implementation of our method. Firstly, the RBF meshfree interpolating functions are seen to be accurate only over a relatively narrow domain in the intrinsic parameter space, which leads to using narrow boundaries to carry out the PE analysis. This constraint restricts the flexibility of the meshfree PE analysis compared to standard approaches. A potential solution to overcome this issue could be to divide the intrinsic parameter space into smaller “patches” and generate interpolants for each patch independently [196].

In the present implementation, the prior bounds (albeit narrow) were obtained from the best-matched template as explained in Sec. 4.3.1. Since the PE pipeline is expected to be triggered by an upstream search pipeline, we have used the most significant trigger parameters ($\vec{\lambda}^*$) as the starting point. Using a suitable optimization algorithm, this point ultimately leads us to a nearby reference point ($\vec{\lambda}_{\text{ref}}$), which serves as the center of the hyper-rectangle from which the samples are drawn. The dimensions of this hyper-rectangle along the different directions are heuristically set to get good accuracy in the likelihood reconstruction (i.e. we do not use the LVK results apriori to determine the prior bounds). In future, we would like to explore an automated way of determining the dimensions of this hyper-rectangle from the size of the 90% overlap ellipsoid calculated at $\vec{\lambda}_{\text{ref}}$. Such an idea has been mooted in Pankow et al. [197]. However, since the axes of the overlap ellipsoid do not necessarily align with the eigen-directions of the data-driven covariance matrix, such an implementation would require careful thought. Secondly, it is worth noting that despite improvements in the placement of nodes compared to our previous work, it still needs investigation for a more generic prescription that can be applied to any CBC system. Incorporating an adaptive strategy [198] to identify significant nodes could enhance interpolation accuracy and strengthen the node placement strategy.

In this analysis, we have focused exclusively on non-precessing signal models (`TaylorF2` and `IMRPhenomD`), with a particular emphasis on the dominant harmonic of the GW signal. These models do not incorporate tidal terms [199, 200]. Since the tidal deformability parameters corresponding to the binary components are intrinsic in nature, we need to include them as

interpolating parameters in our method, which increases the dimensionality of the intrinsic parameter space to six. Additionally, a greater number of nodes are required to ensure an accurate interpolation of the SVD coefficients and template norm square. We will explore these ideas in future follow-ups of this work (see Appendix 4.6.4 for preliminary results).

Although we anticipate that extending our method to include higher-order harmonics should be relatively straightforward, it may result in increased start-up costs. In the case of precessing-spin waveform models, a meshfree framework is particularly suitable since the number of RBF nodes does not exponentially increase with the dimensionality of the parameter space. In this regard, the relative binning algorithm has been recently implemented for gravitational-wave parameter estimation with higher-order modes and precession [201]. In the future, we would like to overcome these limitations by enhancing the node placement algorithm further, expanding the boundaries of the sample space, and incorporating models that account for in-plane spins [81].

In the next chapter, we apply this coherent PE meshfree method for a large-scale simulation study of the impact of including LIGO-Aundha (A1) in the International Gravitational Wave Detector Network (IGWN) on the localization of BNS sources.

Acknowledgements

We thank Vaibhav Tiwari for carefully going through our manuscript and giving helpful suggestions. We also thank Abhishek Sharma and Sachin Shukla for the helpful discussions at various stages of this work. We especially thank the anonymous referee for their careful review and helpful suggestions.

L. P. is supported by the Research Scholarship Program of Tata Consultancy Services (TCS). S. M. acknowledges INSPIRE fellowship by the Department of Science and Technology (DST), India. A. R is supported by the research program of the Netherlands Organisation for Scientific Research (NWO). A. S. gratefully acknowledges the generous grant provided by the Department of Science and Technology, India, through the DST-ICPS (Interdisciplinary Cyber Physical Systems) cluster project funding. We thank the HPC support staff at IIT Gandhinagar for their help and cooperation.

Code availability: Codes used in this analysis are publicly available in the following Github repository [202].

4.6 Appendix

4.6.1 Effect of different choices of Gaussian distributed nodes on the posteriors

The distribution of input nodes plays an important role in the meshfree PE algorithm as described in Section 4.3.1. One would like to place nodes intelligently so as to provide good coverage near the support of the log-likelihood function, which falls very steeply around its mode. Here, we show the effect of node placement on the posteriors obtained. We choose a certain fraction of RBF nodes (N_{Gauss}) from a normal (Gaussian) distribution described by the covariance matrix. The remaining nodes are distributed uniformly over the sample space. We find that the overall accuracy increases as we choose a higher fraction of Gaussian distributed nodes (particularly on the ι and d_L parameters). Although, the estimations of α , δ , and \mathcal{M} are quite robust against different choices of N_{Gauss} .

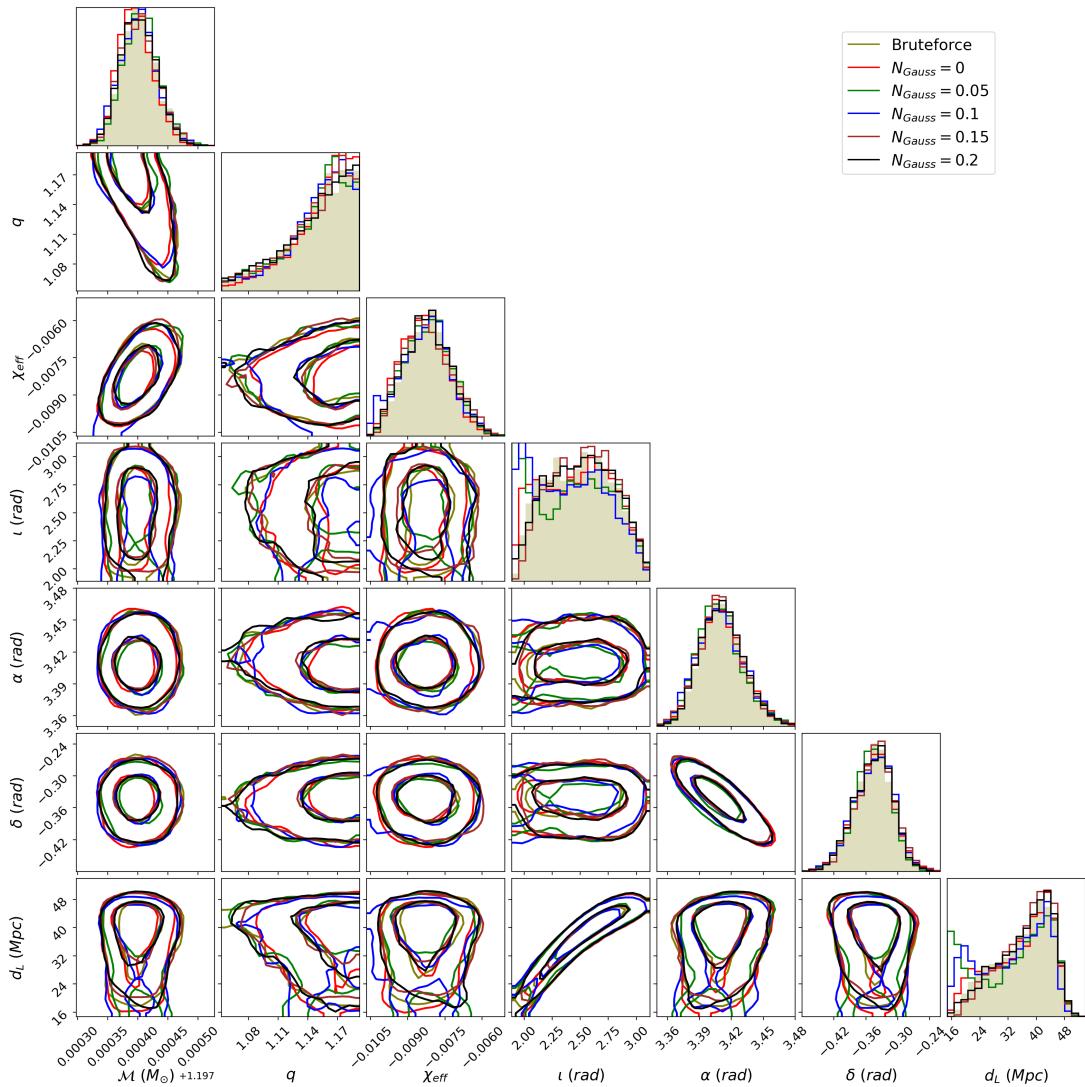


Figure 4.8: The effect of choosing different values of N_{Gauss} (fraction of Gaussian nodes) on the reconstructed parameters.

4.6.2 Likelihood errors

To assess the accuracy of the meshfree method to approximate the likelihood, we generated 10^4 random query points from the prior distribution specified in Table 4.1. Then we calculated the likelihood using both the meshfree and bruteforce likelihood methods. The median absolute error for both TaylorF2 and IMRPhenomD waveform models were found to be $\mathcal{O}(10^{-2})$.

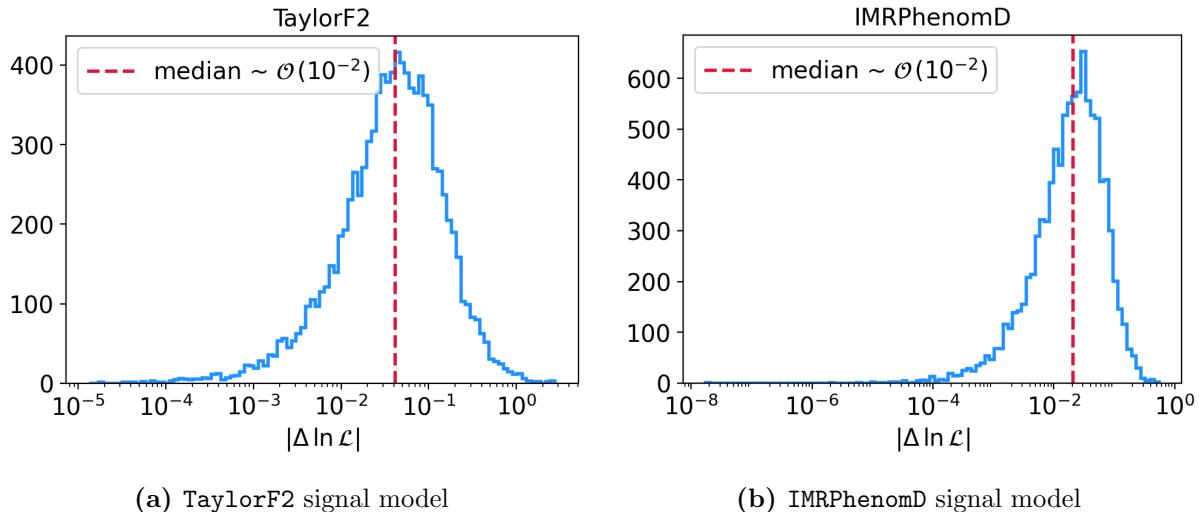


Figure 4.9: The likelihood errors ($\Delta \ln \mathcal{L} \equiv \ln \mathcal{L}_{\text{PyCBC}} - \ln \mathcal{L}_{\text{RBF}}$) are shown for both TaylorF2 and IMRPhenomD waveform models (on GW170817 event). The red-dashed line represents the median absolute error ($|\Delta \ln \mathcal{L}|$) that are $\sim \mathcal{O}(10^{-2})$. There are very few points with high errors ($|\cdot| \sim 4(2)$ for TaylorF2(IMRPhenomD)) which are not shown here.

4.6.3 Surface of SVD coefficients and template norm square

As mentioned in Sec. 4.3.1, the SVD coefficients need to smooth over the intrinsic parameter space in order to be interpolated over it. Fig. 4.10 shows the smooth variation of the SVD coefficients ($C_{\mu}^{q(i)}$) with \mathcal{M} and η . Similarly, the surface of the template norm square ($\sigma^2(\vec{\lambda}^q)^{(i)}$) shows smooth variation over the sample space so that we can interpolate it using RBFs.

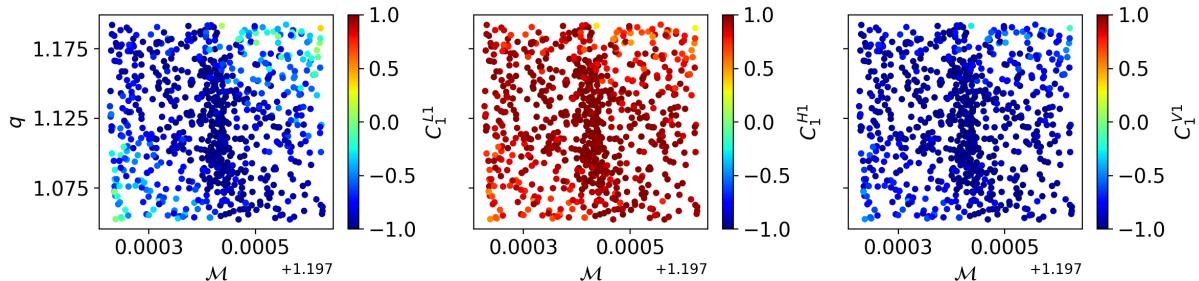


Figure 4.10: Surface of the SVD coefficients corresponding to the first basis for three detectors (L1, H1, and V1). Note that they are smoothly varying functions of \mathcal{M} and η , thereby making them suitable for interpolation.

4.6.4 Addition of tidal deformability in the meshfree framework

To demonstrate meshfree PE for a case where we also include tidal parameters (Λ_1, Λ_2), we use the example of the GW170817 event with the IMRPhenomD_NRTidalv2 [203] waveform model. We choose a uniform prior over Λ_1 and Λ_2 lying between $[100 - 400]$ and spray 2500 nodes, and retain the top-20 basis vectors to be used for snr reconstruction. For comparison, we also run a bruteforce PE (using PyCBC standard likelihood). The meshfree PE took ~ 5.2 minutes (including start-up), whereas the bruteforce PE took ~ 3 hours 25 minutes. Both methods utilized 64 CPU cores. Here are the corner plots of the resulting posterior distributions:

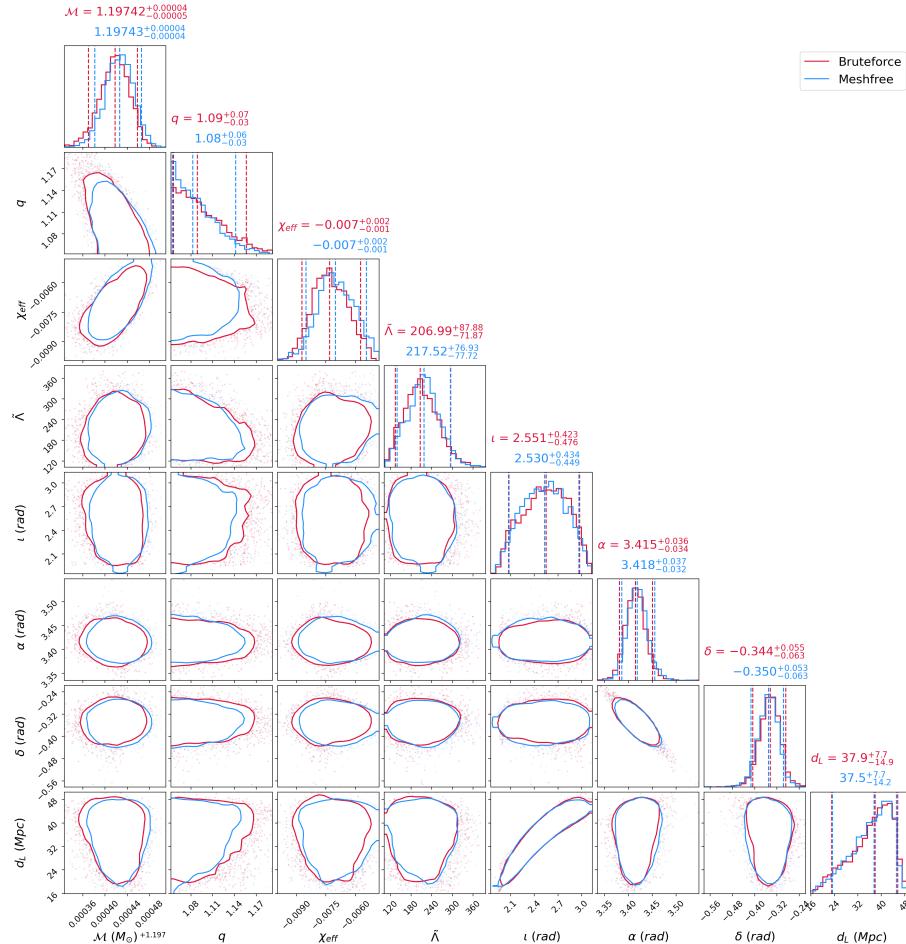


Figure 4.11: Posterior distribution of various binary parameters, including the tidal deformability parameters. Note that we have shown the posterior over $\tilde{\Lambda}$ [14, 204] (a weighted linear combination of two tidal parameters, $\tilde{\Lambda} = \frac{16}{13} \frac{(m_1 + 12m_2)m_1^4\Lambda_1 + (m_2 + 12m_1)m_2^4\Lambda_2}{(m_1 + m_2)^5}$), which is a leading tidal contribution to the GW phase evolution.

4.6.5 GW200115: An NSBH event

We present the results of reconstructing the GW200115 NSBH event using the meshfree and the bruteforce likelihood (for reference). The seismic cutoff frequency is set to 20 Hz, assuming the IMRPhenomD waveform model. We used a 64 seconds data segment around the GW200115 NSBH event, and PSD was generated using the same data with `median-mean` estimation. The center of the sample space was chosen as the MAP values of the PE samples obtained by previous LIGO analysis [205]. The parameters for RBF interpolation were the same as the one used for the GW170817 BNS event earlier, except $\nu = 10$ and $\epsilon = 20$. The sampling parameters for `dynesty` nested sampler were the following: (i) `nlive` = 2500, (ii) `walks` = 350, (iii) `sample` = ‘rwalk’, and (iv) `dlogz` = 0.1. The meshfree likelihood-based PE process (see Fig. 4.12) was completed in ~ 14 minutes when performed on 64 cores.

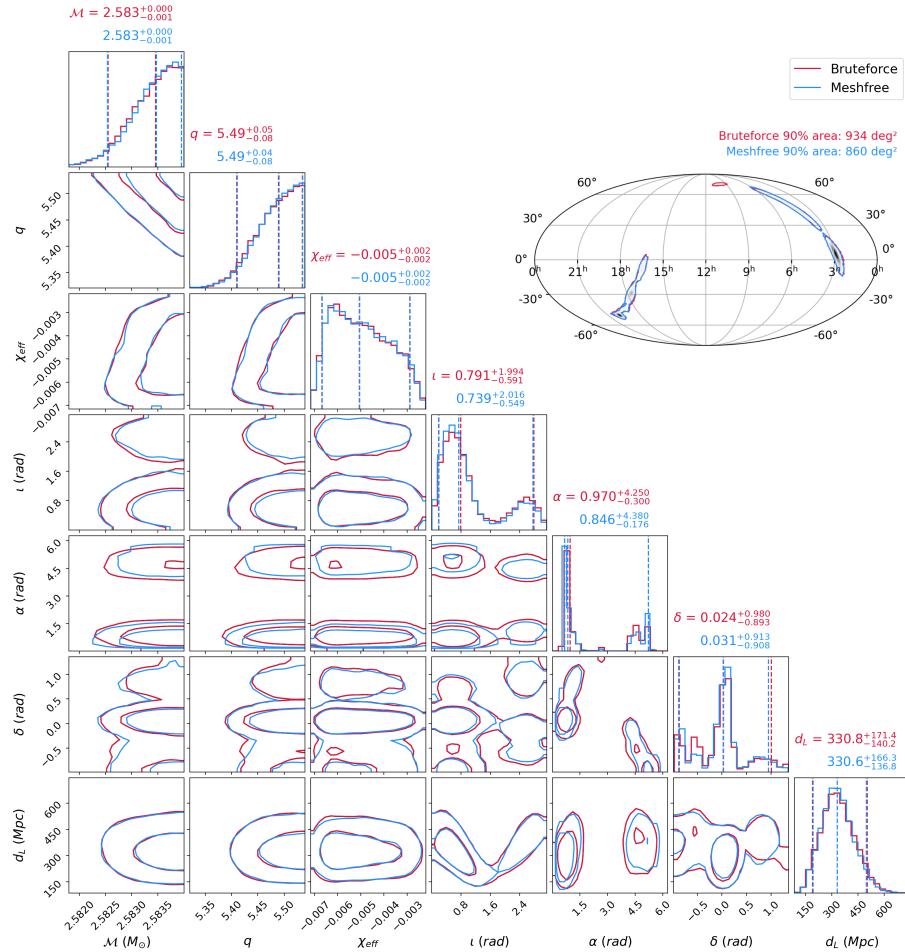


Figure 4.12: The corner plot shows the posterior distributions of ten-dimensional parameters using two different methods: the meshfree method and the bruteforce likelihood. The estimated distributions from both methods exhibit good agreement, as indicated by the median values in the marginalized posterior titles. The *inset* shows the corresponding skymaps and are consistent with the results published by the LIGO, Virgo, and Kagra collaborations [206].

Chapter 5

Pinpointing coalescing binary neutron star sources with the IGWN, including LIGO-Aundha

Abstract LIGO-Aundha (A1), the Indian gravitational wave detector, is expected to join the International Gravitational-Wave Observatory Network (IGWN) and begin operations in the early 2030s. We study the impact of this additional detector on the accuracy of determining the direction of incoming transient signals from coalescing binary neutron star (BNS) sources with moderately high signal-to-noise ratios. It is conceivable that A1’s sensitivity, effective bandwidth, and duty cycle will improve incrementally through multiple detector commissioning rounds to achieve the desired ‘LIGO-A+’ design sensitivity. For this purpose, we examine A1 under two distinct noise power spectral densities. One mirrors the conditions during the fourth science run (O4) of the LIGO Hanford and Livingston detectors, simulating an early commissioning stage, while the other represents the A+ design sensitivity. We consider various duty cycles of A1 at the sensitivities mentioned above for a comprehensive analysis. We show that even at the O4 sensitivity with a modest 20% duty cycle, A1’s addition to the IGWN leads to a 15% reduction in median sky-localization errors ($\Delta\Omega_{90\%}$) to 5.6 sq. deg. At its design sensitivity and 80% duty cycle, this error shrinks further to 2.4 sq. deg, with 84% sources localized within a nominal error box of 10 sq. deg! This remarkable level of accuracy in pinpointing sources will have a positive impact on Gravitational Wave (GW) astronomy and cosmology. Even in the worst-case scenario, where signals are sub-threshold in A1, we demonstrate its critical role in reducing the localization uncertainties of the BNS source. Our results are obtained from a large Bayesian parameter estimation study using simulated signals injected in a heterogeneous network of detectors using the recently developed meshfree approximation aided rapid Bayesian inference pipeline. We consider a seismic cut-off frequency of 10 Hz for all the detectors. We also present hypothetical improvements in sky localization for a few Gravitational-Wave Transient Catalog-like events injected in real data after including a hypothetical A1 detector to the sub-network in which such events were originally detected. We also demonstrate that A1’s inclusion could resolve the degeneracy between the luminosity distance and inclination angle parameters, even in scenarios

where A1 does not directly contribute to improving the network signal-to-noise ratio for the event.

5.1 Introduction

This chapter is based on the publication *Pinpointing Binary Neutron Star sources with a global network of GW Detectors, including LIGO-Aundha*, Phys. Rev. D **109**, 044051 (2024).

The detection and prompt localization of GW170817 event [207] can be regarded as a monumental discovery that led to follow-up observations over the entire electromagnetic (EM) spectrum. This discovery resulted in the first extensive multi-messenger astronomical observing campaign [208] undertaken to follow up post-merger emissions from compact binary coalescence. The concurrent observations of such events via electromagnetic, neutrino, and gravitational wave (GW) detectors facilitate complementary measurements of phenomena that would otherwise remain inaccessible when observed independently. For example, the standard siren measurement of Hubble-Lemaître constant H_0 from the gravitational wave data is independent of the cosmic distance-ladder method as opposed to that in astronomy with EM radiation. This provides a complementary measurement of H_0 , which can be used to elucidate the Hubble-Lemaître tension associated with the disparity between the measurements of Hubble constant [209, 210, 206] in the early and late universe.

Early observations of post-merger emissions of binary neutron star (BNS) sources can be used to constrain the physical models behind the internal mechanisms of these emission processes. For instance, there are different models proposed explaining the origin of the early blue emission of the kilonova associated with GW170817. Even though the model proposing radioactive decay of heavy elements in low-opacity ejecta being a theoretically motivated candidate [211, 212, 213, 214] fits the rise time curve well enough, there are also other models (for example: by cooling of shock-heated ejecta) which fit the decline of the emission equally well. In the case of GW170817, the associated kilonova was detected ~ 11 hours after the merger, limiting the information about the rise time of the kilonova, particularly in the ultraviolet band [211]. Capturing the rise time of the emission light curves by early detection may provide a useful measure to differentiate between the kilonova origin models. The gamma-ray burst (GRB) GRB 170817A was detected independently ~ 1.7 seconds after the trigger time of GW170817, with studies later confirming its association with the BNS merger [215]. This association confirmed BNS mergers as the progenitors of at least certain short GRBs [215]. The simultaneous detection of GWs and GRB may provide remarkable insights into the central engine of short gamma-ray bursts (SGRBs). The time delay between the GW and GRB events (~ 1.7 secs as in the case of GW170817) offers valuable information into the underlying physics and intrinsic processes within the core, including the formation of a remnant object and the subsequent jet. It is expected that the EM waves and GW must have identical propagation speeds. The time delay between the GW and GRB events can be used to put constraints on the deviation of the speed of gravity waves from the speed of light, hence allowing for the tests of fundamental laws of physics [215]. Radio emissions enable

tracing the fast-moving ejecta from BNS coalescence, providing insights into explosion energetics, ejecta geometry, and the merger environment [216, 217]. Meanwhile, X-ray observations are crucial for determining the energy outflow geometry and system orientation relative to the observer’s line of sight [218]. The post-merger transients possess observational time scales ranging from a few seconds to weeks, owing to their potential to harness radiation across the entire EM spectrum. Detecting EM counterparts to ~ 50 BNS mergers may enable determining H_0 with 2% fractional uncertainty [219], sufficient to verify the presence of any systematic errors in the local measurements of H_0 . Even in the absence of potential EM counterparts, the accurate localization of compact binary coalescence (CBC) sources allows for dark siren measurements of H_0 [220, 221, 222]. For instance, observations from more than ~ 50 BNS dark sirens may be required to obtain H_0 measurements with 6% fractional error [219]. Thus, pinpointing these mergers is key in providing promising probes of fundamental physics, astrophysics, and cosmology.

However, in order to accurately locate the EM counterparts and supplement regular follow-ups, more accurate and rapid 3D-localization of the sources using GW observations is of utmost importance. With imminent improvements in the detector sensitivities for future observing runs, there is an inevitable need for rapid Parameter Estimation (PE) methods. This arises from the fact that an increased bandwidth of detectors towards the lower cut-off frequencies would result in a momentous increase in the computational cost of Bayesian PE, hence affecting the prompt localization of the GW source. Currently, LIGO-Virgo-KAGRA Collaboration (LVK) [95] uses a Bayesian, non-MCMC-based rapid sky localization tool, known as **BAYESTAR** [223], to locate (posterior distributions over sky location parameters, α , and δ) CBC sources within tens of seconds following the detection of the corresponding GW signal. However, as shown by Finstad et al. [129], a full Bayesian analysis including both intrinsic and extrinsic parameters can significantly increase the accuracy of sky-localization (by $\sim 14 \text{ deg}^2$ in their analysis) of the CBC sources. This is of primary significance in reducing the telescope survey area & time for locating the EM counterparts, making rapid Bayesian PE-based methods a preferable as well an evident necessity. Nevertheless, since **BAYESTAR** can construct skymaps in the order of a few tens of seconds, it would be an interesting case to test if the **BAYESTAR** skymap samples can be used as priors for sky location estimates in rapid Bayesian PE methods. This scheme might enable a more accurate localization measurement of the source. Future improvements in detector noise sensitivities shall lead to an increase in the detector “*sensemion range*”¹ [226, 225, 227]. Hence, higher BNS detections are expected from distances further away than current ranges [228]. It is projected that ~ 180 BNS events could be detected in future O5 run [229]. Hence, it is judicious to start the EM follow-up right from the merger epoch, which would require early warning alerts [230, 231] for the EM telescopes in future runs. Given the finite observational resources of EM facilities allotted for the GW localized regions, it is important to prioritize the follow-up based on the chirp-mass estimates of GW events, as suggested by Margalit et

¹“*sensemion range*” is defined as the radius of a sphere of volume in which a GW detector could detect a source at a fixed SNR threshold, averaged over all sky locations and orientations [224]. For lower redshifts ($z \lesssim 1$), the *sensemion range* is approximately equal to the horizon distance divided by 2.264. [225]

al. [130]. This can be facilitated by a rapid Bayesian PE analysis, which, in addition to the sky localization, more importantly, provides a significantly accurate estimation of chirp mass for these compact binary systems. However, there are studies [232] which show that the chirp mass can be estimated accurately (with uncertainty no larger than $\sim 10^{-3} M_{\odot}$) in low-latency searches. Although the mass ratio and effective aligned spin estimates may be severely biased in this case. With the addition of new detectors in the ground-based detector network, it becomes important to observe improvements expected in the sky localization and source parameter estimations of these compact binary sources.

LIGO-Aundha [47, 48, 49, 50] (hereafter, ‘A1’) is set to join the network of ground-based GW detectors in the early years of the next decade. Based on our experience with currently operational interferometric GW detectors, we expect A1 to go through multiple commissioning rounds, resulting in incremental improvement to its noise sensitivity, effective bandwidth, and duty cycle. We expect that it will eventually attain the target `aLIGO` A+ design sensitivity [95, 233] (sometimes referred to as the ‘O5’ sensitivity in this paper) and acquire higher up-times, resulting in duty cycles that are commensurate with the stable operations of other detectors in the IGWN. To model this progression, we consider A1 to have two distinct noise sensitivities: one at conditions that mirror the fourth science run (O4) of the LIGO Hanford and Livingston detectors, simulating an early commissioning stage, and the other at the A+ design sensitivity of the detector. Additionally, we consider various duty cycles of the A1 detector at the sensitivities mentioned above for a comprehensive analysis.

By the time A1 begins its maiden science run, the current detectors namely LIGO-Hanford (H1) and LIGO-Livingston (L1) [41], are likely to reach their A+ design sensitivities or beyond; meanwhile Virgo (V1) [159] and KAGRA (K1) [234] are also expected to reach their target design sensitivities. A1 would add significantly longer baselines to the existing GW network and also contribute to increasing the network SNR [235, 236], leading to better sky localizations of GW sources along with improvements in the estimations of source parameters. A number of studies done previously have addressed the localization capabilities of GW networks which include analytical studies [237, 238] as well as simulations using `BAYESTAR`, `LaInference` or other localization algorithms [239, 240, 228, 241, 242]. In this article, we aim to provide an illustration of the contribution A1 shall make to the current network of terrestrial GW detectors with a focus on improvements in sky localization capabilities for BNS mergers. We focus on the scenarios where the localization uncertainties are of the orders that enable a potential EM follow-up. This is possible when the source is localized by three or more detectors. We perform a full Bayesian Parameter Estimation using a rapid PE method developed by Pathak et al. [243, 148]. The method enables us to perform rapid Bayesian PE for BNS systems from a lower seismic cut-off frequency $f_{\text{low}} = 10 \text{ Hz}$, hence allowing for the bandwidths (especially at lower frequencies) that would be typical of the future (O5 & beyond) observing runs. Analyzing CBCs from $f_{\text{low}} = 10 \text{ Hz}$ increases the number of cycles in the frequency band and leads to an improvement in signal-to-noise ratio (SNR) accumulation and information content of the event.

The study of the localization capabilities of a GW detector network can be broadly characterized by the intrinsic source parameters (e.g. component masses) [239, 228] of binary

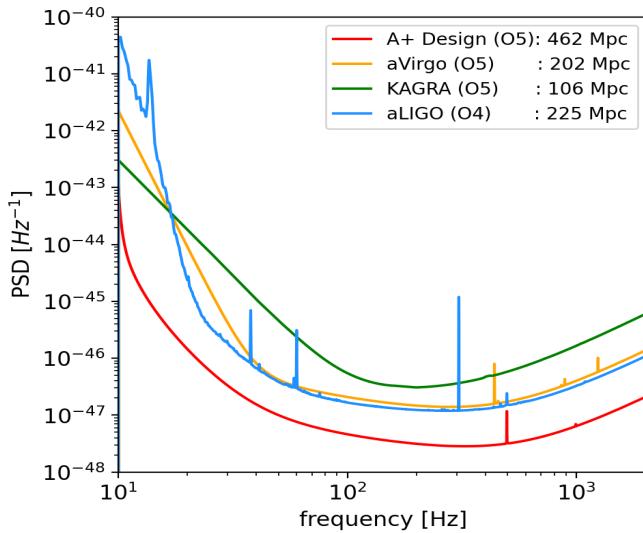


Figure 5.1: The noise Power Spectral Density (PSD) for GW detectors. L1 and H1 are at aLIGO A+ Design Sensitivity(O5). V1 and K1 are at aVirgo(O5) and KAGRA(80 Mpc) (O5) sensitivities respectively. We study A1 at two different sensitivities: first at aLIGO (O4) and then at aLIGO A+ Design Sensitivity (O5), respectively. The distances quoted here are the “sensemon” ranges for the respective detectors for a $1.4M_{\odot} + 1.4M_{\odot}$ BNS at $\rho_{\text{th}} = 6$. Note that the sensemon ranges for these PSD estimates presented in [95, 233] correspond to a detection signal-to-noise ratio threshold of 8 in a single detector.

systems, detector sensitivities, and duty cycles of individual detectors [235, 228]. We hereby discuss these aspects in relation to our work.

5.1.1 Detector Networks

The ground-based GW detector network is comprised of detectors with different sensitivities. The heterogeneous nature of the noise Power Spectral Density (PSD) curves of detectors plays an important role in deciding the localization abilities of the network. Hence, we assume the ground-based GW detectors to be at different noise sensitivities for our analysis. The first two LIGO detectors L1 and H1 are configured to aLIGO A+ Design Sensitivity PSDs [41, 95]. The detector V1 is taken to be at its projected aVirgo PSD [159]. Meanwhile, the K1 detector is assumed to be at KAGRA (80 Mpc) O5 design sensitivity [234, 44, 43].

To study the improvements in localization capabilities of the network with the addition of A1, we analyze the A1 detector at two different sensitivities. We first consider the case where A1 is in its initial operating phase at aLIGO O4 sensitivity. In the second case, we take A1 to be at aLIGO A+ Design Sensitivity configuration, marking the target sensitivity it shall achieve after undergoing staged commissioning over time. All the above-mentioned PSDs can be found in [233]. The aforementioned noise sensitivity curves are shown in Fig. 5.1. These sensitivity configurations, in conjunction with the duty cycles taken into account, shall provide a more comprehensive approach in presenting a science case for the addition of A1 to the current GW detector network.

5.1.2 Duty Cycles

The duty cycle of a detector/network is defined as the fraction of time for which it successfully collects data of scientific significance during an observing run [236]. The duty cycle for a detector depends on the specific phase in which the detector is relative to its intended target operating configuration. Additionally, environmental effects also play a role in affecting the detector duty cycle during an observation run. No detector can practically acquire science-quality data at all times. This translates to detectors working at different duty cycles depending on the commissioning of the detectors.

To understand the impact of duty cycles on the localization of BNS sources by a network, we *assume* the following three cases as suggested by Pankow et al. [228] representing different stages of a detector's operation:

- 1) 20% duty cycle: A representative of the early stages of commissioning and engineering runs, resulting in reduced operational time.
- 2) 50% duty cycle: A representative of unresolved technical issues with the instrumental setup as well as signaling challenges like suboptimal environmental conditions.
- 3) 80% duty cycle: A representative of a detector operating near the possible target operating point.

Consider a GW network with N detectors. Over the course of an observing run, there can be a k number of detectors ($N_{\min} \leq k \leq N$) participating in data collection, depending on their duty cycles. Here, N_{\min} represents the minimum number of detectors assumed to be participating in the observation of a GW event. These k participating detectors can comprise different subnetworks of distinct detectors. For instance, a GW network with $N = 5$ detectors may have only $k = 4$ detectors in operation. Subject to which detector is out of operation, there can be $\binom{N}{k} = 5$ different subnetworks of distinct detectors. Depending on the duty cycles of individual detectors, the effective duty cycle of a subnetwork can be evaluated. Out of a total of N detectors in the network, we assume a set \mathbf{m} composed of all the detectors participating in data collection and a set \mathbf{n} comprising detectors that are out of operation (possibly due to maintenance) during this observation period. We represent the probability (p) of being in operation defined by a given duty cycle (i.e. $p = 0.5$ for 50% duty cycle). The probability representing the effective duty cycle of a subnetwork (p_{eff}) is given as

$$p_{\text{eff}} = \prod_{m_i} \prod_{n_j} p_{m_i} (1 - p_{n_j}), \quad (5.1)$$

where p_{m_i} and p_{n_j} are the duty cycles of the i th detector in set \mathbf{m} and j th detector in set \mathbf{n} respectively. For instance, consider a network of $N = 4$ detectors, namely L1, H1, V1, and K1. The probability of being in operation for individual detectors is given by p_{L1} , p_{H1} , p_{V1} , and p_{K1} representing their respective duty cycles. There may be a scenario where any one of these four detectors may get out of operation due to maintenance or environmental causes. Depending on which detector is not in observation mode, there can be $\binom{4}{3} = 4$ subnetworks, comprising of $k = 3$ detectors participating in data acquisition. To evaluate the probability of being in operation p_{eff}

for one of the subnetworks consisting of say H1, V1, and K1 detectors (this is the case when L1 is not in operation) we have

$$p_{\text{eff}} \Big|_{\text{H1V1K1}} = p_{\text{H1}} \cdot p_{\text{V1}} \cdot p_{\text{K1}} \cdot (1 - p_{\text{L1}})$$

where, $\text{H1}, \text{V1}, \text{K1} \in \mathbf{m}$ and $\text{L1} \in \mathbf{n}$ respectively. The effective duty cycle for N -detector network is evaluated as the probability obtained by adding the probabilities of being in operation for all subnetworks over $k \leq N$ number of participating detectors.

In this study, we consider the L1, H1, V1, and K1 detectors operating near their target operating point at 80% duty cycle each and are fixed during the analysis. As A1 is expected to join this network by the early 2030s, we aim to show how the addition of A1 improves the localization capabilities of the GW network. We vary duty cycles for the A1 detector, simulating the cases for various phases of its configuration relative to its target operating point. Single detectors are nearly omnidirectional due to the structure of the antenna pattern functions - leading to poor source localization. For a two-detector network, solving for the direction in the sky corresponding to a fixed time-delay between the coalescence times recorded at the detectors leads to a ‘ring’ like pattern on the sky, and as such, events localized with two detectors are generally not useful enough for EM follow-ups. Hence, we do not include the cases with $k \leq 2$ in our analysis. For the purpose of this study, we assume $3 \leq k \leq N$ (i.e. $N_{\min} = 3$) and evaluate the network duty cycles accordingly. The impact of A1 as an addition to the second-generation GW network consisting of L1, H1, V1, and K1 detectors is presented by the implementation of varying duty cycles (20%, 50%, 80%) in conjunction with different detector sensitivities (**aLIGO O4** and **aLIGO A+ Design Sensitivity**) for A1 detector.

5.1.3 Injected BNS sources

The remarkably high signal-to-noise ratio (SNR) due to the fortunate proximity of the GW170817 event enabled its effective localization and multi-messenger efficacies. Since the number of such ‘golden events’ is expected to be low even in future observing runs [235, 244], it becomes increasingly important to study a network’s ability to localize such events. Hence, we aim to focus on studying the impact of the addition of the LIGO-Aundha detector in localizing moderately high SNR events, which may lead to potential multi-messenger observations. Taking this into consideration, we choose to generate 500 BNS events having an optimal network SNR in the range of 20 to 25 in the GW network comprising L1, H1, V1, and K1 detectors. We use these events for the purpose of our investigation. We would like to highlight that this is not a population study but focuses on possible improvements to the localization capabilities of a GW network with the addition of the LIGO-Aundha detector in accurately locating such ‘golden events’. An event is considered to be detected if the individual detector optimal SNR is greater than a threshold value of 6 ($\rho_{\text{th}} > 6$) in at least two detectors. From all the generated BNS sources, an event must follow the detection criteria to be considered detected by a subnetwork/network. The effectiveness of a network with an additional A1 detector is studied against the four-detector network with L1, H1, V1, and K1 detectors.

The intrinsic parameters, like component masses, spins, etc., affect the localization of CBC

sources. The effective bandwidth, as defined in [237], measures the frequency content of the signal. Effective bandwidth is one of the important factors affecting the localization of CBC sources [239]. The signals from BNS sources mostly span through the entire bandwidth of the ground-based detectors owing to their relatively small component mass values in comparison to binary black holes or neutron star-black hole sources. The mass ranges for the BNS are also narrow, leading to small variations in effective bandwidths. In fact, it has also been shown by Pankow et al. [228] that the sky localization uncertainties for BNS systems are effectively independent of the population model of their component masses and spins. Hence, in order to simplify our simulations, we work with a particular choice of component source masses and spins. The tidal parameters are also expected to have a negligible effect on the source localization of these systems [239] and, therefore, are not included as source parameters.

The source-frame intrinsic parameters are chosen to have the maximum a posteriori (MAP) values of the posterior samples obtained from the LIGO PE analysis of GW170817 [245] using `BILBY` [246] Python package. The source frame component masses are $m_1^{\text{src}} = 1.387 M_{\odot}$, $m_2^{\text{src}} = 1.326 M_{\odot}$, while the dimensionless aligned spin component parameters are $\chi_{1z} \approx 1.29 \times 10^{-4}$, $\chi_{2z} \approx 3.54 \times 10^{-5}$. We choose the inclination angle $\iota = \pi/6$ radians. We set the polarization angle arbitrarily to $\psi = 0$ radians. The sources are distributed uniformly in sky directions. We distribute the sources in luminosity distances corresponding to the redshifts following a uniform in comoving volume distribution up to a redshift of ~ 0.14 , which is greater than the detection range of a detector at **aLIGO A+ Design Sensitivity** (O5) for a BNS with component masses m_1^{src} and m_2^{src} . In addition to the sources that are to be detected with certainty, this limit also allows for events that are barely near the threshold in some detectors. We simulate uncorrelated Gaussian noise in each detector characterized by their associated PSDs respectively. The injection sources are generated with the `IMRPhenomD` [102] waveform model, and the source parameters are recovered using the `TaylorF2` [98, 99, 100, 101] waveform model for the Bayesian PE analysis. Since BNS mergers are mostly inspiral-dominated in the LIGO-Virgo-KAGRA (LVK) detector band, the use of the `TaylorF2` waveform model sufficiently extracts the required information from strain data.

5.1.4 Bayesian inference

In order to estimate the parameters of GW sources, we use a Bayesian framework, where for a given waveform model \mathbf{h} and data \mathbf{d} from the detectors, the posterior distribution of the source parameters $\vec{\Lambda}$ can be estimated via Bayes' theorem:

$$p(\vec{\Lambda} | \mathbf{d}) = \frac{\mathcal{L}(\mathbf{d} | \vec{\Lambda}) p(\vec{\Lambda})}{p(\mathbf{d})} \quad (5.2)$$

where $\mathcal{L}(\mathbf{d} | \vec{\Lambda})$ is the likelihood function, $p(\vec{\Lambda})$ is the prior over the source parameters $\vec{\Lambda} \equiv \{\vec{\lambda}, \vec{\theta}, t_c\}$, and $p(\mathbf{d})$ is called evidence, which describes the probability of data given the model. Here $\vec{\lambda}$ represents the intrinsic parameters, whereas $\vec{\theta}$ denotes the extrinsic parameters. In principle, $p(\vec{\Lambda} | \mathbf{d})$ can be estimated by placing a grid over the parameter space $\vec{\Lambda}$, which for a typical compact binary coalescence (CBC) source described by a ~ 15 dimensional parameter

space, would become practically intractable. In the case of a BNS system, it increases to 17 dimensional space due to the addition of two tidal deformability parameters. Instead, stochastic sampling methods such as Markov chain Monte Carlo (MCMC) [143] and Nested Sampling [71] are employed to generate representative samples from the posterior distribution $p(\vec{\Lambda} | \mathbf{d})$. However, this process still requires evaluating the likelihood function, which involves a computationally expensive step of generating model (template) waveforms at the proposed points by the sampler and calculating the overlap between these waveforms and the data. This computational cost is notably significant, especially for low-mass systems such as binary neutron star (BNS) events with lower cutoff frequency decreased to 10 Hz. The situation is exacerbated by the enhanced sensitivity of detectors, resulting in a large number of in-band waveform cycles. Furthermore, incorporating additional physical effects can further escalate the computational burden of waveform generation. These factors have significant implications for the feasibility of promptly following up on EM counterparts of corresponding BNS systems. As mentioned earlier, a high number of BNS detections are expected in the O5 runs; it would be prudent to prioritize the EM follow-ups given the limited observational resources. This underscores the importance of the development of rapid PE methods, which can efficiently estimate both intrinsic and extrinsic parameters. Various rapid PE methods have been proposed in the recent past. They broadly come under two categories: (i) “likelihood-based” approaches such as Reduced order models [111, 112, 170, 113, 106], Heterodyning (or Relative Binning) [139, 138, 129, 168], and other techniques such as RIFT [137], simple-pe [171], multibanding [124] (ii) “likelihood-free” approaches which aim to directly learn the posteriors employing Machine-learning techniques such as deep learning, normalizing flows, and variational inference as well [116, 136, 175, 176, 132].

In this work, we use a likelihood-based rapid PE method developed by Pathak et al. [243, 148], which combines dimensionality reduction techniques and meshfree approximations to swiftly calculate the likelihood at the proposed query points by the sampler. This algorithm is interfaced with DYNESTY [72, 73], a Python implementation of the Nested sampling algorithm to quickly estimate the posteriors distribution over the source parameters. In the forthcoming sections, we will first define the likelihood function and subsequently provide a concise overview of how the meshfree method expeditiously computes the likelihood at the sampler’s proposed query points.

Likelihood function

Given a stream of data $d^{(i)}$ from the i^{th} detector and a template $\tilde{h}^{(i)}(\vec{\Lambda})$, under an assumption of uncorrelated noise across the detectors, the coherent network log-likelihood is given by

$$\ln \mathcal{L}(\vec{\Lambda}) = \sum_{i=1}^{N_d} \langle \mathbf{d}^{(i)} | \tilde{h}^{(i)}(\vec{\Lambda}) \rangle - \frac{1}{2} \sum_{i=1}^{N_d} \left[\|\tilde{h}^{(i)}(\vec{\Lambda})\|^2 + \|\mathbf{d}^{(i)}\|^2 \right] \quad (5.3)$$

where $\tilde{h}^{(i)}(\vec{\Lambda})$ represents the frequency domain Fourier Transform (FT) of the signal $h^{(i)}(\vec{\Lambda})$ and N_d is the number of detectors. Here, the inner product is defined as

$$\langle x | y \rangle = 4 \operatorname{Re} \int_0^\infty df \frac{\tilde{x}(f)^* \tilde{y}(f)}{S_h(f)} \quad (5.4)$$

In this paper, we focus on the non-precessing GW signal model, which can be decomposed into factors dependent on only intrinsic and extrinsic parameters as follows:

$$\begin{aligned}\tilde{h}^{(i)}(\vec{\Lambda}) &\equiv \tilde{h}(\vec{\Lambda}, t^{(i)}) = \mathcal{A}^{(i)} \tilde{h}_+(\vec{\lambda}, t^{(i)}), \\ &= \mathcal{A}^{(i)} \tilde{h}_+(\vec{\lambda}) e^{-j 2\pi f t_c} e^{-j 2\pi f \Delta t^{(i)}}\end{aligned}\quad (5.5)$$

where $\mathcal{A}^{(i)}$, the complex magnitude of the signal depends only on the extrinsic parameters $\vec{\theta} \in \vec{\Lambda}$ through the antenna pattern functions, luminosity distance d_L , and the inclination angle ι , and can be expressed as the following:

$$\mathcal{A}^{(i)} = \frac{1}{d_L} \left[\frac{1 + \cos^2 \iota}{2} F_+^{(i)}(\alpha, \delta, \psi) - j \cos \iota F_\times^{(i)}(\alpha, \delta, \psi) \right], \quad (5.6)$$

$\Delta t^{(i)}$ corresponds to the time-delay introduced due to the relative positioning of the i^{th} detector in relation to the Earth's center [150], the $F_+^{(i)}(\alpha, \delta, \psi)$ and $F_\times^{(i)}(\alpha, \delta, \psi)$ are respectively the ‘plus’ and ‘cross’ antenna pattern functions of the i^{th} detector, which are functions of right-ascension α , declination δ , and polarization angle ψ . The antenna pattern functions describe the angular response of the detector to incoming GW signals [183].

In our analysis, we opt for the log-likelihood function marginalized over the coalescence phase [145]. With $\tilde{h}^{(i)}(\vec{\Lambda})$ given by Eq. (5.5), the expression of the marginalized phase likelihood is given by

$$\begin{aligned}\ln \mathcal{L}(\vec{\Lambda} \mid \mathbf{d}^{(i)}) \Big|_{\phi_c} &= \ln I_0 \left[\left| \sum_{i=1}^{N_d} \mathcal{A}^{(i)*} \langle \mathbf{d}^{(i)} \mid \tilde{h}_+(\vec{\lambda}, t^{(i)}) \rangle \right| \right] \\ &\quad - \frac{1}{2} \sum_{i=1}^{N_d} \left[\left| \mathcal{A}^{(i)} \right|^2 \sigma^2(\vec{\lambda})^{(i)} + \|\mathbf{d}^{(i)}\|^2 \right];\end{aligned}\quad (5.7)$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind and $\tilde{z}^{(i)}(\vec{\lambda}^n) \equiv \langle \mathbf{d}^{(i)} \mid \tilde{h}_+(\vec{\lambda}, t^{(i)}) \rangle$ is the complex overlap integral, while $\sigma^2(\vec{\lambda})^{(i)} \equiv \langle \tilde{h}_+(\vec{\lambda}, t^{(i)}) \mid \tilde{h}_+(\vec{\lambda}, t^{(i)}) \rangle$ is the squared norm of the template $\tilde{h}_+(\vec{\lambda})$. $\sigma^2(\vec{\lambda})^{(i)}$ depends on the noise power spectral density (PSD) of the i^{th} detector. The squared norm of the data vector, $\|\mathbf{d}^{(i)}\|^2$, remains constant throughout the PE analysis and, hence, does not affect the overall ‘shape’ of the likelihood. Consequently, it can be excluded in the subsequent analysis. Note that the marginalized phase likelihood will not be an appropriate choice for systems with high precession and systems containing significant power in subdominant modes [247].

Meshfree likelihood interpolation

The meshfree likelihood interpolation, as outlined in [243, 148], comprises two stages: (i) Start-up stage, where we generate radial basis functions (RBF) interpolants of the relevant quantities and (ii) Online-stage, where the likelihood is calculated by evaluating the interpolants at the query points proposed by the sampler. Let’s briefly discuss both stages.

- **Start-up stage:** First, we generate N RBF interpolation nodes in the intrinsic parameter space (\mathcal{M} , q , χ_{1z} , and χ_{2z} in this context). The center $\vec{\lambda}^{\text{cent}}$ around which these interpolation

nodes are positioned is determined by optimizing the network-matched filter SNR, starting from the best-matched template or trigger $\vec{\lambda}^{\text{trig}}$ and t_{trig} identified by the upstream search pipelines [141, 248, 15]. For simulated systems, the injection parameter is taken as the central point for node placement. We employ a combination of Gaussian and uniform nodes, where the Gaussian nodes are sampled from a multivariate Gaussian distribution (MVN) with a mean of $\vec{\lambda}^{\text{cent}}$ and a covariance matrix calculated using the inverse of the Fisher matrix evaluated at $\vec{\lambda}^{\text{cent}}$. A hybrid node placement approach ensures that nodes are positioned near the peak of the posterior, where higher accuracy in likelihood reconstruction is necessary. Once the nodes $\vec{\lambda}^n$ are generated, we efficiently compute the time-series $\vec{z}^{(i)}(\vec{\lambda}^n) \equiv z^{(i)}(\vec{\lambda}^n, t_c)$ using the Fast Fourier Transform (FFT) circular correlations, with t_c being uniformly spaced discrete-time shifts within a specified range (± 150 ms²) around a reference coalescence time t_{trig} . During this calculation, we set $\Delta t^{(i)} = 0$ for overlap time series, handling extra time offsets introduced due to sampling in the sky location parameters during the online stage. Similarly, we compute the template norm square $\sigma^2(\vec{\lambda}^n)^{(i)}$ at the RBF nodes $\vec{\lambda}^n$. We then stack the time series (row-wise) and perform Singular Value Decomposition (SVD) of the resulting matrix, producing a set of basis vectors spanning the space of $\vec{z}^{(i)}(\vec{\lambda}^n)$:

$$\vec{z}^{(i)}(\vec{\lambda}^n) = \sum_{\mu=1}^N C_{\mu}^{n(i)} \vec{u}_{\mu}^{(i)} \quad (5.8)$$

where the SVD coefficients $C_{\mu}^{n(i)}$, smooth functions of $\vec{\lambda}^n$ within the sufficiently narrow boundaries encompassing the posterior support, can be interpolated over the $\vec{\lambda}$ using a linear combination of RBFs and monomials [107]:

$$C_{\mu}^{q(i)} = \sum_{n=1}^N a_n^{(i)} \phi(\|\vec{\lambda}^q - \vec{\lambda}^n\|_2) + \sum_{j=1}^M b_j^{(i)} p_j(\vec{\lambda}^q) \quad (5.9)$$

where ϕ is the RBF kernel centered at $\vec{\lambda}^n \in \mathcal{R}^d$, and $\{p_j\}$ denotes the monomials that span the space of polynomials with a predetermined degree ν in d -dimensions. Since the coefficients are only known at N RBF nodes $\vec{\lambda}^n$, we impose M additional conditions of the form $\sum_{j=1}^M a_j^{(i)} p_j(\vec{\lambda}^q) = 0$ to uniquely solve for the coefficients a_n and b_j in the Eq. (5.9). Furthermore, it turns out that only “top-few” basis vectors are sufficient to reconstruct $\vec{z}^{(i)}(\vec{\lambda}^q)$ at minimal reconstruction error. Consequently, we generate only top- ℓ meshfree interpolants of $C_{\mu}^{q(i)}$ where $\mu = 1, \dots, \ell$, where ℓ can be chosen based on the singular value profile. Similarly, we express $\sigma^2(\vec{\lambda}^q)^{(i)}$ in terms of RBFs and monomials, treating them as smoothly varying functions over the interpolation domain. Finally, we have uniquely constructed the $\ell + 1$ RBF interpolants, which are to be used in the online stage.

- **Online stage:** In the online stage, we rapidly compute interpolated values of $C_{\mu}^{q(i)}$ and $\sigma^2(\vec{\lambda}^q)^{(i)}$ at any query point $\vec{\lambda}^q$ within the interpolation domain. Subsequently, we determine

²This range should be larger than the maximum light travel time between two detectors.

the corresponding $\vec{z}^{(i)}(\vec{\lambda}^q)$ using Eq.(5.8). Rather than generating the entire time series, we focus on creating $\vec{z}^{(i)}(\vec{\lambda}^q)$ with around $\mathcal{O}(10)$ time samples centered around the query time $t^{q(i)}$, which contain the additional time-offset $\Delta t^{(i)}$. We fit these samples with a cubic spline, from which we calculate $\vec{z}^{(i)}(\vec{\lambda}^q)$ at the query time $t^{q(i)}$. Similarly, we compute the interpolated value of $\sigma^2(\vec{\lambda}^q)^{(i)}$. Finally, we integrate these interpolated values with the factors related to extrinsic parameters, as outlined in Eq.(5.7), to compute the interpolated likelihood $\ln \mathcal{L}_{\text{RBF}}$.

Parameters	Range	Prior distribution
\mathcal{M}_{det}	$[\mathcal{M}_{\text{det}}^{\text{cent}} \pm 0.0001]$	$\propto \mathcal{M}_{\text{det}}$
q	$[q^{\text{cent}} \pm 0.07]$	$\propto [(1+q)/q^3]^{2/5}$
$\chi_{1z,2z}$	$[\chi_{1z}^{\text{cent}} \pm 0.0025]$	Uniform
V_{com}	$[\sim 16.9e3, \sim 8.9e8]$	Uniform
t_c	$t_{\text{trig}} \pm 0.12$	Uniform
α	$[0, 2\pi]$	Uniform
δ	$\pm\pi/2$	$\sin^{-1} [\text{Uniform}[-1, 1]]$
ι	$[0, \pi]$	Uniform in $\cos \iota$
ψ	$[0, 2\pi]$	Uniform angle

Table 5.1: Prior parameter space over the ten-dimensional parameter space $\vec{\Lambda}$.

5.2 Analysis of Simulated Events

As discussed previously in Section 5.1.3, we create injections with fixed source-frame masses and dimensionless aligned spin component parameters. However, the detector-frame parameters (masses) for these events vary according to their associated redshifts. We define the intrinsic detector-frame parameters by $\vec{\lambda} = (\mathcal{M}_{\text{det}}, q, \chi_{1z}, \chi_{2z})$. Similarly, the injected intrinsic parameters in the detector frame are denoted as $\vec{\lambda}^{\text{cent}}$. To perform Bayesian PE for each event, we first generate $N_{\text{nodes}} = 800$ RBF nodes as described in Section 5.1.4. We sample 20% of the total RBF nodes ($N_{\text{Gauss}} = 160$) from a multivariate Gaussian distribution $\mathcal{N}(\vec{\lambda}^{\text{cent}}, \boldsymbol{\Sigma})$, where $\vec{\lambda}^{\text{cent}}$ is the mean and $\boldsymbol{\Sigma}$ is the covariance matrix obtained from inverse of the Fisher matrix $\boldsymbol{\Gamma}$ around the center $\vec{\lambda}^{\text{cent}}$ using the `gwfast` [190] python package. The remaining 80% of the total RBF nodes ($N_{\text{Unif}} = 640$) are sampled uniformly from the ranges provided in Table 5.1. We choose $\phi = \exp(-(\epsilon r)^2)$ as the Gaussian RBF kernel in our analysis, with ϵ being the shape parameter. For the purpose of this analysis, we use $\epsilon = 10$, monomial terms with degree $\nu = 7$ and $l = 20$ top basis vectors for reconstructing the time-series in Eq. (5.8). After the successful generation of interpolants, the likelihood function can be evaluated using $\ln \mathcal{L}_{\text{RBF}}$ by sampling the ten-dimensional parameter space $\vec{\lambda}$ using the `dynesty` sampler. The sampler configuration

is outlined as follows: `nlive` = 500, `walks` = 100, `sample` = “rwalk”, and `dlogz` = 0.1. These parameters play a critical role in determining both the accuracy and the time required for the nested sampling algorithm to converge. In this context, the parameter `nlive` represents the number of live points. Opting for a larger value of `nlive` leads to a more finely sampled posterior distribution (and consequently, the evidence), but it comes at the cost of requiring more iterations to achieve convergence. The parameter `walks` specifies the minimum number of points necessary before proposing a new live point, `sample` indicates the chosen approach for generating samples, and `dlogz` represents the proportion of the remaining prior volume’s contribution to the total evidence. In this analysis, `dlogz` = 0.1 serves as a stopping criterion for terminating the sampling process. For a more comprehensive understanding of `dynesty`’s nested sampling algorithm and its practical implementation, one can refer to the following references [72, 73].

The prior distribution for $\vec{\lambda}$, along with the associated parameter space boundaries, are presented in Table 5.1. The prior distributions for the extrinsic parameters $(\alpha, \delta, V_{\text{com}}, \iota, \psi, t_c)$ and their respective parameter space boundaries are also presented in Table 5.1. To evaluate the Bayesian posteriors of source parameters, we sample over the entire ten-dimensional parameter space involving four intrinsic and six extrinsic source parameters. This ensures accounting for the correlations between parameters. However, the focus of this study lies in discussing the sky localization uncertainties obtained from the posteriors over α and δ parameters.

In accordance with the previous discussion in Section 5.1.4, we perform PE for the simulated events with different subnetworks of a GW network to take into account the effect of duty cycles. For instance, in the case of a network with L1, H1, V1, K1, and A1, there can be 10 different subnetworks consisting of three distinct detectors ($k = 3$), and 5 different subnetworks of four distinct detectors ($k = 4$) taking observations depending on the duty cycles. In addition to these, there is a subnetwork consisting of all the five detectors for $k = 5$ case.

Bayesian PE analyses are performed for the events detected in each of these subnetworks. The total number of subnetworks for all $3 \leq k \leq 5$ is 16 for the five detector networks comprising L1, H1, V1, K1, and A1 detectors. The exercise is repeated for two cases:

- (i) Keeping A1 at `aLIGO` O4 noise sensitivity in the GW network. Here, the A1 sensitivity is close to `aVirgo` (O5) sensitivity (Refer Fig. 5.1).
- (ii) Setting A1 at `aLIGO A+ Design Sensitivity` (O5) in the GW network. In this case, the A1 detector would be at the same sensitivity as the other two LIGO detectors.

We represent the network with $N = 5$ detectors as the L1H1V1K1A1 network and, similarly, the network with $N = 4$ detectors as the L1H1V1K1 network. Using the `ligo-skymap` [223] utility, we compute the 90% credible sky localization areas $\Delta\Omega_{90\%}$ (in sq. deg) from the posterior samples over right ascension (α) and declination (δ) obtained from Bayesian PE.

5.3 Sky Localization Results

In order to take into account the effect of duty cycles in the sky localization of our simulated BNS events, we first evaluate the probabilities associated with the effective duty cycles of each

subnetwork of a detector network using Eq. (5.1). Each subnetwork is assigned a fixed number of events depending on their observation probabilities. Taking this into account and integrating all such cases across $3 \leq k \leq N$ with area samples of corresponding events gives the localization $\Delta\Omega_{90\%}$ distribution related to the network duty cycle for a given GW network.

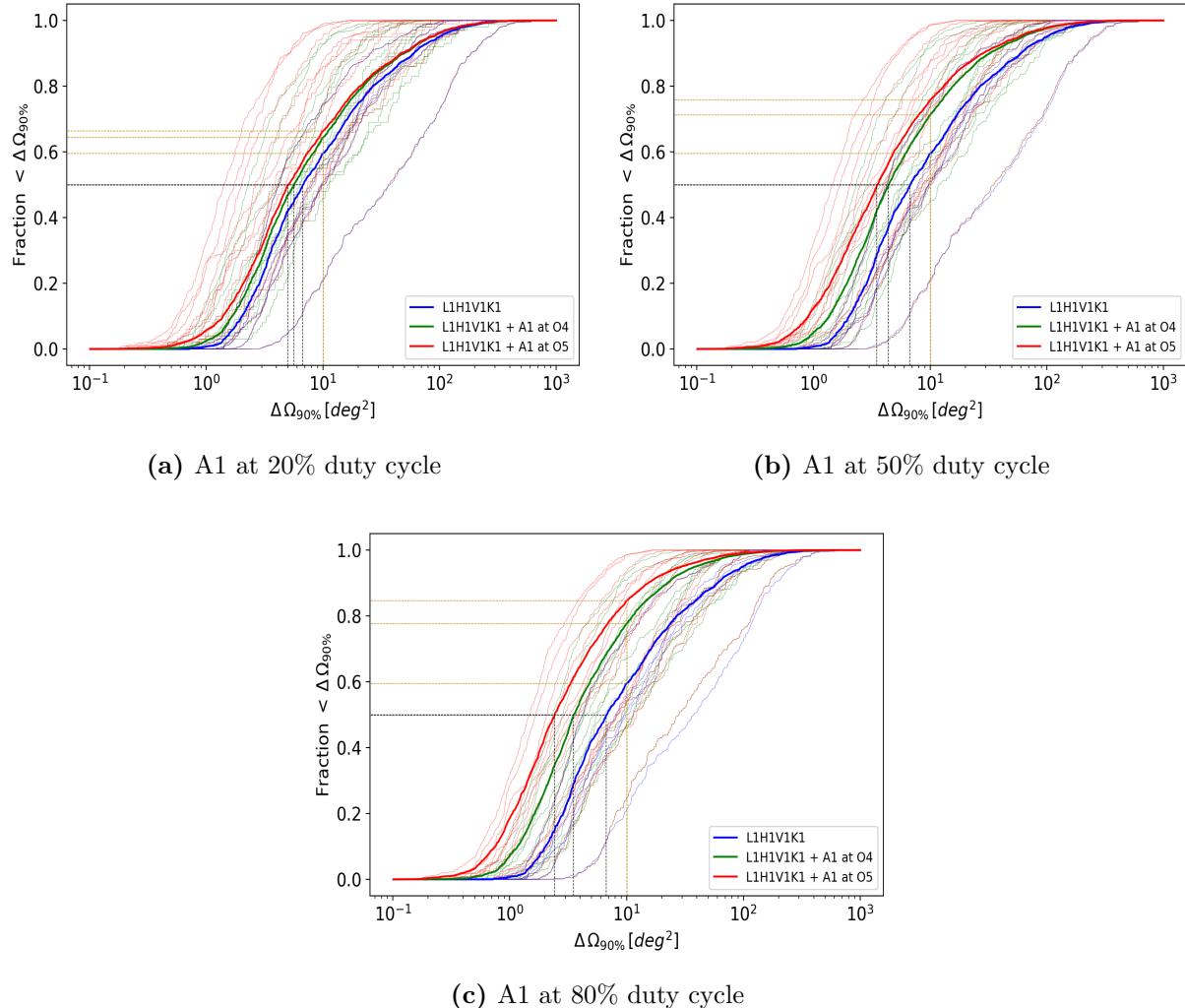


Figure 5.2: Cumulative Distributions (CDF) of 90% localization area of simulated BNS events obtained by different networks are shown. As mentioned in 5.1.2, duty cycles for each of the L1, H1, V1, and K1 detectors are set to 80%. Three different cases considering A1 at duty cycles of 20%, 50%, and 80% are shown, along with taking into account two different noise sensitivities (A1 at O4 sensitivity and A1 at O5 sensitivity) and presented in Fig. 5.2a, 5.2b and 5.2c respectively. The blue solid lines represent the CDF obtained by the L1H1V1K1 network with the aforementioned duty cycle. The green solid lines represent the CDF constructed by the L1H1V1K1A1 network, where A1 is at `aLIGO` (O4) sensitivity. The red solid lines are the CDF constructed with L1H1V1K1A1 network, where A1 is at `A+ Design Sensitivity` (O5) sensitivity. In each of the subplots, the A1 detector is at a duty cycle, as mentioned in the captions of Fig. 5.2a, 5.2b and 5.2c respectively. The lighter curves represent the CDFs of $\Delta\Omega_{90\%}$ areas from different subnetworks over all possible k for the corresponding GW network. The vertical black color dashed lines mark the median localization areas obtained by each network. The horizontal golden color dashed lines represent the fraction of events (in percentage) recovered under 10 sq. deg. by each detector network.

Network	Fraction within < 10 sq. deg.(in %) & Median $\Delta\Omega_{90\%}$ Area (sq. deg.)		
L1H1V1K1	59%		
(Median $\Delta\Omega_{90\%}$)	6.6		
	A1 at 20% duty cycle	A1 at 50% duty cycle	A1 at 80% duty cycle
L1H1V1K1+A1 (O4)	64%	71%	77%
(Median $\Delta\Omega_{90\%}$)	5.6	4.3	3.5
L1H1V1K1+A1 (O5)	66%	75%	84%
(Median $\Delta\Omega_{90\%}$)	4.9	3.4	2.4

Table 5.2: We present the percentage of detected events localized within 10 sq. deg sky area and the median $\Delta\Omega_{90\%}$ for different detector networks. The median $\Delta\Omega_{90\%}$ for detector networks are shown separately in the rows below the values representing the percentage of detected events localized within 10 sq. deg sky area. The noise sensitivity of A1, along with the duty cycles associated with A1, are mentioned in the relevant table sections. Duty cycles for L1, H1, V1, and K1 detectors are fixed at 80%.

The results of our simulations are presented in Fig. 5.2. The Cumulative Distribution Function (CDF) plots in Fig. 5.2 are constructed from $\Delta\Omega_{90\%}$ sky area samples obtained by inverse sampling from the localization distributions for each subnetwork. We find that with the L1H1V1K1 network, the median 90% localization area $\Delta\Omega_{90\%}$ is 6.6 sq. deg., meanwhile 59% of the BNS sources are localized within less than 10 sq. deg. area in the sky. With the addition of an A1 detector to this network, we find significant improvements in the localization capabilities of the terrestrial detector network. Similar improvements were also reported by [235, 236]. It is also evident from Fig. 5.2 that duty cycles and detector noise sensitivities play a vital role in the effective localization of sources. We shall discuss these in detail as follows:

5.3.1 A1 at aLIGO-O4 sensitivity

As a part of the five-detector network, the A1 detector is initially set to aLIGO O4 noise sensitivity. The median $\Delta\Omega_{90\%}$ area in the decreasing order are found to be 5.6, 4.3, and 3.5 in sq. deg. when A1 is set to 20%, 50%, and 80% duty cycle respectively. We find that 64%, 71%, and 77% of the events are localized with less than 10 sq. deg in sky area, given that A1 is at 20%, 50%, and 80% duty cycles respectively. Our results suggest that even when A1 is at 20% duty cycle, which can be interpreted as the early commissioning phase of the detector, the five-detector network reduces the median $\Delta\Omega_{90\%}$ localization uncertainty to 5.6 sq. deg. in comparison to 6.6 sq. deg. obtained by the four detectors L1H1V1K1 network. This reduction in the sky localization area plays a crucial role in the ‘tiled mode’ search for EM counterparts undertaken by the EM facilities such as the GROWTH India Telescope [249] with a field of view of 0.38 sq. deg. in area, to tile the GW localization regions. As the A1 detector is upgraded to be at 80% duty cycle, the median localization area $\Delta\Omega_{90\%}$ remarkably reduces by approximately a factor of two in comparison to that achieved by the four detector L1H1V1K1 network. We reiterate that

the L1, H1, V1, and K1 detectors are taken to be operating at 80% duty cycles.

5.3.2 A1 at A+ Design (O5) sensitivity

By upgrading the A1 configuration to aLIGO A+ Design Sensitivity (O5), the improvement in the localization capabilities of the five-detector network relative to the four-detector network as well as the five-detector network with A1 at O4 sensitivity is considerable. The median $\Delta\Omega_{90\%}$ localization uncertainties in the decreasing order are found to be 4.9, 3.4, and 2.4 sq. deg. in area, when A1 is set to 20%, 50% and 80% duty cycle respectively. We find that 66%, 75%, and 84% of the events are localized with less than 10 sq. deg in sky area, given that A1 is at 20%, 50% and 80% duty cycles respectively, where A1 is set at A+ sensitivity. We observe that with the A1 detector operating at 50% duty cycle, the median localization area $\Delta\Omega_{90\%}$ reduces by a factor of two with respect to the values obtained by the L1H1V1K1 network. As A1 reaches its target operating point with 80% duty cycle, we find the median $\Delta\Omega_{90\%}$ to reduce by a factor of three against the median localization achieved with the four-detector network. As mentioned previously, for A1 operating at O4 sensitivity and 80% duty cycle, the median $\Delta\Omega_{90\%}$ is 3.5 sq. deg., whereas this reduces to 2.4 sq. deg. when A1 is set to 80% duty cycle and at O5 sensitivity.

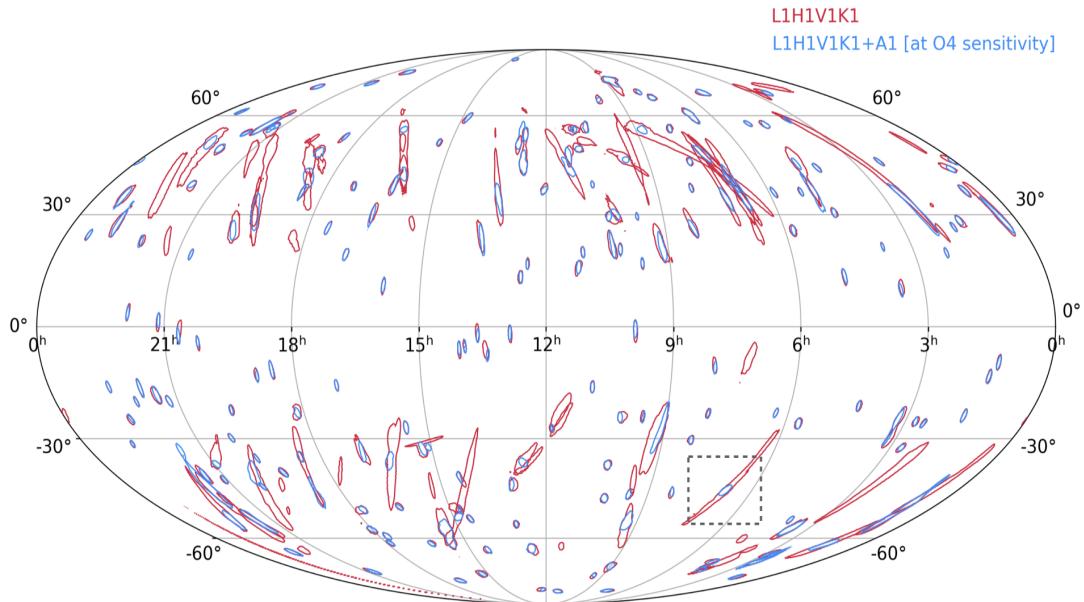


Figure 5.3: The crimson and blue probability contours correspond to 90% credible sky area ($\Delta\Omega_{90\%}$) localization of the events by L1H1V1K1 and L1H1V1K1A1 networks, respectively for 191 events out of our 500 simulated BNS events. Here, A1 is at O4 sensitivity, and the events are subthreshold in A1. The contours marked in a dashed square box are discussed in Section 5.4. This skymap plot exhibits the improvement achieved in the localization of BNS events by the inclusion of an A1 detector, even when it does not contribute to the detection. Note that for some sky locations, the red contours appear in the background of the blue contours and may not be visible.

The A1 detector, with an upgraded O5 sensitivity, leads to a two-fold impact on the network capabilities. On the one hand, it leads to an increase in the number of detections (events satisfying the SNR threshold $\rho_{\text{th}} \geq 6$) in the subnetworks, including A1. It also results in an increase in network SNR, which is one of the important factors contributing to the effective localization of sources. By adding A1 to the GW network, there is an increase in the observation probability of three or more detectors by 7% (for A1 operating at the early 50% duty cycle) in comparison to the four detector networks. We summarize a few important remarks about the localization $\Delta\Omega_{90\%}$ for the different network configurations in Table 5.2. Note that the CDF plots in Fig. 5.2 may indicate slightly lower median $\Delta\Omega_{90\%}$ localization values than those obtained in related studies [228, 239]. One of the reasons is that the events are analyzed from $f_{\text{low}} = 10$ Hz, which increases the effective bandwidths, as well as due to the exclusion of cases with two detectors subnetworks or single detectors participation in source localization.

The focus of this study is to explore the localization capabilities of the GW network with A1, with possibilities leading to potential EM follow-ups and providing better ground for astrophysical and cosmological investigations.

5.4 Localization of simulated BNS events subthreshold in A1

During the observation of GW170817, the event was detected in L1 and H1 detectors but was below the detection threshold in the V1 detector. Yet, the presence of V1 contributed to localizing the source to a few tens of sq. deg. Out of the 500 simulated BNS events described in our previous discussion, a total of 191 events detected in the five-detector network were found to be subthreshold ($\rho_{\text{th}} < 6$) in A1 is at **aLIGO** O4 sensitivity. We compare the sky-localizations for these events obtained from a four-detector L1H1V1K1 network to those achieved by the L1H1V1K1A1 network. Since these events are subthreshold in A1, the contribution of A1 in improving the network SNR is negligible. Yet, the presence of an A1 detector leads to an improvement in reducing the localization uncertainties of these events. This is shown in Fig. 5.3. We find that even in the case where these events are subthreshold in A1 (at **aLIGO** O4 sensitivity), the percentage of events localized with less than 10 sq. deg. in the sky increase from 72% to 89% in comparison to the L1H1V1K1 network. Even though the CDFs (for $\Delta\Omega_{90\%}$) used for the estimation of these improvements are not very smooth due to a lesser number of such events (191 in this case), they summarize the essence of overall nature of the improvement well enough.

As the noise sensitivity configuration of A1 is upgraded to **aLIGO A+ Design Sensitivity**, there is an increase in the number of detections in the A1 detector. In this case, the number of events that are subthreshold in the A1 detector reduces from 191 to just 44 out of all the 500 simulated BNS sources. Due to the improved sensitivity, further improvements in the localization of such events are achieved in comparison to the localizations obtained relative to the four-detector L1H1V1K1 network and L1H1V1K1A1 network with A1 at O4 sensitivity. Note that in the context of this section, we do not consider the duty cycles for these networks. Therefore, we make a direct comparison between the localization results achieved for such events with the L1H1V1K1 network and the L1H1V1K1A1 network. For instance, the event marked in

Fig. 5.3 is localized to 44 sq. deg. with L1H1V1K1 network. The same event when detected by the L1H1V1K1A1 network with A1 is at **aLIGO O4** sensitivity, is recorded at an optimal SNR value $\rho_{A1} = 3.1$ in A1 detector ($\rho_{A1} < \rho_{\text{th}}$) and is localized to ~ 6 sq. deg. Meanwhile, when A1 is set to **aLIGO A+ Design Sensitivity**, this event is localized to ~ 3.5 sq. deg. area in the sky. The baselines added to the network with the addition of the A1 detector, and its antenna patterns are some of the factors leading to better localization of such events. An improved noise PSD results in an increase in effective bandwidth and hence leads to the reduction in localization uncertainties.

5.5 Experiments with GWTC-like events in real noise

In the preceding section, we showed that even if a detector does not detect an event, it nevertheless adds a valuable contribution to the network in localizing the source. In this section, we provide an illustration of how the incorporation of an additional detector could have facilitated the source localization of events from GWTC for compact binary mergers. The two BNS events, GW170817 and GW190425, and an NSBH event, GW200115, are chosen as examples from GWTC for this purpose. In our analysis, we consider A1 as a supplementary detector. We simulate the aforementioned events and inject them into real detector noise to account for a realistic scenario. The noise strains from the L1, H1, and V1 detectors were acquired by using the **GWpy** [250] python package, which allows the extraction of noise strain timeseries from the datasets publicly available on GWOSC [251]. The noise strain for A1 is taken to be that of the detector, which recorded the least SNR during the observation of these events. In fact, among all the detectors observing these events, the lowest individual SNR was recorded in the Virgo detector. The noise strain data for all the detectors is chosen hundreds of seconds away from the trigger times of the GW events in consideration. Even though the A1 noise strain is taken from V1 data, the noise strain data for both detectors belong to different stretches of data. The PE analysis for these events is performed from a lower seismic frequency of 20 Hz.

5.5.1 Noise Strain and PSDs

A noise strain data of a fixed segment length (360 seconds for GW170817 and GW190425; 64 seconds for GW200115) is used in our analysis, which is cleaned by a high-pass filter of 4th order and setting the frequency cut-off at 18 Hz. We analyze the event from a lower seismic cutoff frequency of 20 Hz, which is illustrative of the observing runs associated with their detections. The estimation of noise PSD uses 2 seconds overlapping segments of the strain data with the implementation of the median-mean PSD estimation method from PyCBC [15]. The noise PSD for A1 is constructed from the strain data of the V1 detector.

5.5.2 Choice of injection parameters:

GW170817-like event

The intrinsic detector-frame parameters ($m_1^{\text{det}}, m_2^{\text{det}}, \chi_{1z}, \chi_{2z}$) and extrinsic parameters like inclination (ι) take values chosen by evaluating the MAP values from the posterior samples of detector-frame parameters in the `Original BILBY results file` [245] for GW170817. We assume the BNS system with spins aligned in the direction of orbital angular momentum. The sky location coordinates of NGC 4993—the potential host galaxy of the GW170817 event, are taken as the injection values for (α, δ) sky position parameters [252]. The luminosity distance takes the value $d_L = 40.4$ Mpc [253] for our simulated BNS system. The polarization angle is taken to be zero ($\psi = 0$) since the associated posterior samples in `Original BILBY results file` are found to be degenerate. As mentioned previously, the simulated signal is injected in uncorrelated real noise in the detectors. The simulated signal is generated using the `IMRPhenomD` waveform model. The source parameters are recovered using the `TaylorF2` waveform model (see Fig. 5.6 in Appendix 5.8.1 for the corner plot and skymap).

GW190425-like event

The intrinsic (detector-frame) and extrinsic parameter values are chosen by evaluating the MAP values of the posterior samples for parameters obtained from `C01:IMRPhenomPv2_NRTidal:LowSpin` LIGO analysis file of GW190425 event [254]. The polarization angle is chosen to be $\psi = 0$ as the posterior samples for ψ from the LIGO analysis follow a uniform distribution. We generate the simulated signal using the `IMRPhenomD` waveform model. The source parameters are recovered using the `TaylorF2` waveform model for the PE analysis (see Fig. 5.7 in Appendix 5.8.1 for the corner plot and skymap).

GW200115-like event

For simulating the NSBH event, we choose the intrinsic (detector-frame) and extrinsic parameter values by evaluating the MAP values of the posterior samples for parameters from `C01:IMRPhenomNSBH:LowSpin` file of GW200115 LIGO analysis [255]. We take the polarization angle $\psi = 0$. We generate the simulated signal using the `IMRPhenomD` waveform model. For recovering the source parameters during the PE analysis, we again use the `IMRPhenomD` waveform model, as it also accounts for the post-inspiral regime, which occurs within the LIGO-Virgo band for the NSBH system (see Fig. 5.8 in Appendix 5.8.1 for the corner plot and skymap).

5.5.3 Analysis and Configurations

The prior distributions and prior boundaries for $(\mathcal{M}_{\text{det}}, q, \chi_{1z}, \chi_{2z}, V_{\text{com}})$ parameters, chosen for the three simulated events are presented in Table 5.3. The priors for the parameters $(t_c, \alpha, \delta, \iota, \psi)$ are same as that shown in Table 5.1 and hence are not separately mentioned here. The Bayesian PE analysis follows a similar methodology of generating interpolants for the likelihood function, as discussed previously. The analysis involves the generation of RBF nodes. We specify

the total number of RBF nodes (N_{nodes}) by mentioning the number of nodes sampled from a multivariate Gaussian $\mathcal{N}(\vec{\lambda}^{\text{cent}}, \Sigma)$, represented by N_{Gauss} ; meanwhile the number of nodes uniformly sampled around $\vec{\lambda}^{\text{cent}}$ are represented as N_{Unif} for each event. The `dynesty` sampler configurations are also mentioned in Table 5.3 for the three simulated events.

	GW170817-like	GW190425-like	GW200115-like	
Parameter	Prior Range	Prior Range	Prior Range	Prior Distribution
\mathcal{M}_{det}	$[\mathcal{M}_{\text{det}}^{\text{cent}} \pm 0.0002]$	$[\mathcal{M}_{\text{det}}^{\text{cent}} \pm 0.0005]$	$[\mathcal{M}_{\text{det}}^{\text{cent}} \pm 0.0011]$	$\propto \mathcal{M}_{\text{det}}$
q	$[1, 1.14]$	$[1, 1.28]$	$[q^{\text{cent}} \pm 0.1]$	$\propto [(1 + q)/q^3]^{2/5}$
$\chi_{1z,2z}$	$[\chi_{1z,2z}^{\text{cent}} \pm 0.0025]$	$[\chi_{1z,2z}^{\text{cent}} \pm 0.0025]$	$[\chi_{1z,2z}^{\text{cent}} \pm 0.0025]$	Uniform
V_{com}	$[5e3, 1e8]$	$[5e3, 4e8]$	$[1e6, 3e8]$	Uniform
No. of RBF Nodes	$N_{\text{nodes}} = 800$ $(N_{\text{Gauss}} = 20\%)$	$N_{\text{nodes}} = 800$ $(N_{\text{Unif}} = 100\%)$	$N_{\text{nodes}} = 1100$ $(N_{\text{Gauss}} = 10\%)$	
RBF Parameters	$\epsilon = 10, \nu = 7$	$\epsilon = 10, \nu = 7$	$\epsilon = 30, \nu = 10$	$\phi = \exp(-(\epsilon r)^2)$
Sampler Configurations	<code>nLive=500</code> <code>nwalks=100</code> <code>sample="rwalk"</code> <code>dlogz = 0.1</code>	<code>nLive=1500</code> <code>nwalks=500</code> <code>sample=="rwalk"</code> <code>dlogz = 0.1</code>	<code>nLive=500</code> <code>nwalks=100</code> <code>sample="rwalk"</code> <code>dlogz = 0.1</code>	

Table 5.3: The prior parameter space is presented in the top part of the table. The hybrid placement for the RBF nodes is mentioned by specifying N_{Gauss} and N_{Unif} for each event, followed by the Gaussian RBF kernel parameters. The `dynesty` sampler configurations chosen for the PE analysis of each event are also presented in the end. The priors for the extrinsic source parameters ($t_c, \alpha, \delta, \iota, \psi$) are same as that shown in Table 5.1 and hence are not mentioned here.

The network comprising the L1, H1, and V1 detectors detected the GW170817 event. As discussed earlier, we simulate a GW170817-like signal in the non-Gaussian real noise and find the source localization uncertainty in the presence of an A1 detector added to the L1H1V1 network. The matched-filter SNR in L1, H1, V1 and A1 are 22.6, 18.6, 5.4 and 6.3 respectively for the given noise realization. Even though, in this case, the addition of A1 to the L1H1V1 network does not lead to any considerable improvements in the network matched-filter SNR for the event, a significant reduction in 90% credible localization area is observed. We find the localization uncertainty $\Delta\Omega_{90\%}$ to be 15 sq. deg for the L1H1V1 network, whereas the localization area $\Delta\Omega_{90\%}$ reduces to 6 sq. deg. with L1H1V1A1 network. Hence, the localization uncertainty is reduced by a factor of more than two in this case. The localization probability contours representing $\Delta\Omega_{90\%}$ obtained from the two different networks for GW170817-like event is presented in Fig. 5.4. For a GW190425-like event, we compare the sky localization with the then-observing network of the L1V1 network to the L1V1A1 network. The sky localization uncertainty ($\Delta\Omega_{90\%}$) reduces from 9350 sq. deg. with the L1V1 network to a sky region of area 212 sq. deg with the L1V1A1 network. The matched filter SNR in L1, V1, and A1 are 10.1, 5.1, and 5.3, respectively, for this case. It is evident that the event in V1 and A1 is at

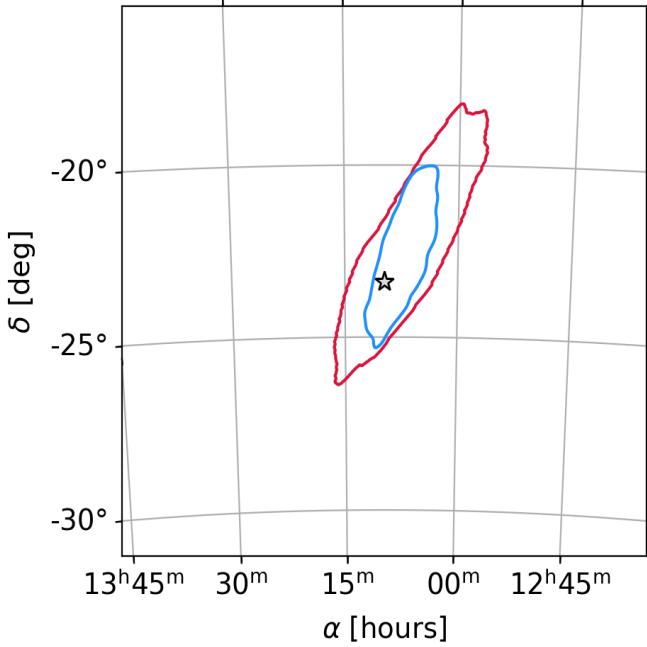


Figure 5.4: The crimson and blue probability contours represent 90% credible localization area ($\Delta\Omega_{90\%}$) for the GW170817-like event by L1H1V1 and L1H1V1A1 networks respectively. Noisy real strain data from V1 are used as a surrogate for A1 in this exercise. The star marks the true location (α, δ) for the source. Though A1 does not contribute to the network SNR, its presence leads to a significant improvement in source localization.

subthreshold SNR for the given noise realization. Yet, there is a contribution in reducing the sky localization areas. For the case of the GW200115-like event, the source localization with the L1H1V1 network, which was the observing network during the real event, is compared to that with the L1H1V1A1 network. The source localization error ($\Delta\Omega_{90\%}$) is reduced from 662 sq. deg. obtained with L1H1V1 to 87 sq. deg. achieved with L1H1V1A1. Here, we observe that the majority of the SNR is accumulated by the initial two LIGO detectors. Meanwhile, V1 and A1 contribute negligibly to improving the network SNR. This is because both V1 and A1 are at similar noise sensitivities for the aforementioned events.

Note that these results vary with different realizations of the detector noise. Nevertheless, the antenna patterns and baselines added to a network by incorporating an additional detector (here, A1) may lead to an enhancement in the localization abilities of the network, even if the signal is subthreshold in one of the detectors.

5.5.4 Degeneracy between luminosity distance and inclination angle

The GW190425-like event, when observed by the L1V1 network, shows a degeneracy between luminosity distance and inclination angle parameters, which was also observed for the real event. As the number of detectors in the network increases from L1V1 to L1V1A1, we observe a resolution of the distance-inclination angle degeneracy. For further investigation, we present the case for GW190425-like events with 28 different real non-Gaussian noise realizations. The

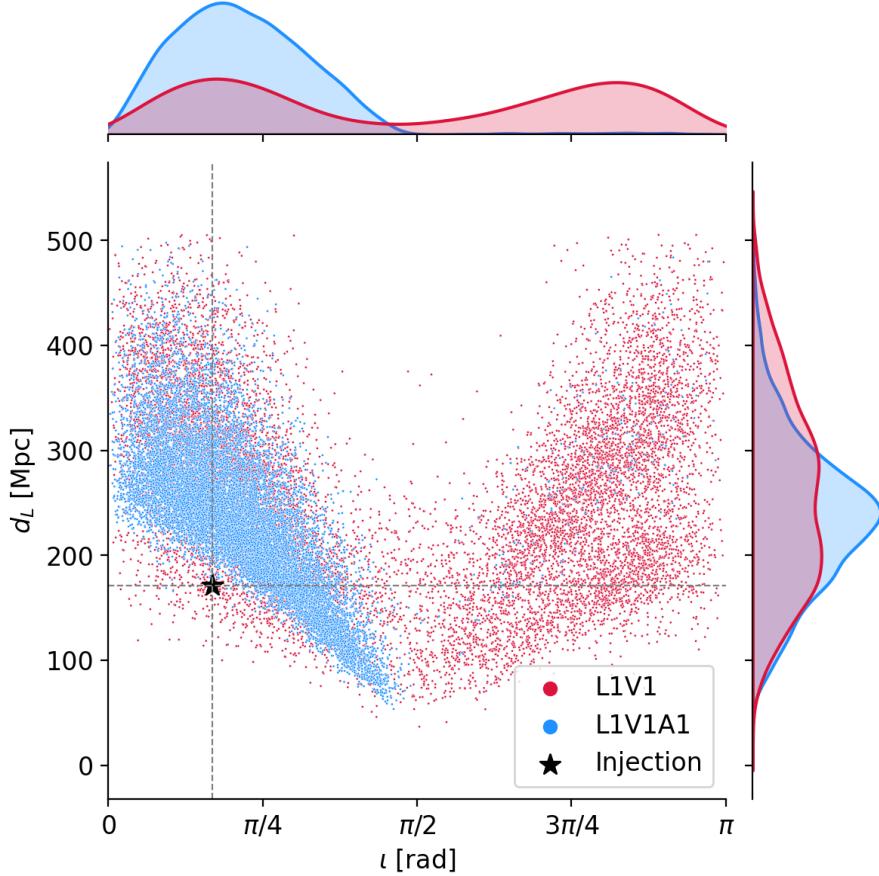


Figure 5.5: The crimson and blue points mark the posterior samples over d_L and ι for the two-detector L1V1 and three-detector L1V1A1 networks respectively. The posteriors are constructed over 28 different noise realizations. This signifies that a three-detector network (L1V1A1) with a third detector contributing negligibly to the network SNR (in this case, A1) is able to play an important role in resolving the $\iota-d_L$ degeneracy relative to a two detector network (L1V1).

noise strains and PSD are obtained as mentioned at the beginning of Section 5.5, where different noise strains correspond to different segments of detector strain data. The events for which the injected chirp-mass (\mathcal{M}_c) is within the 90% credible interval of the posterior samples are chosen.

The results are summarized in Fig. 5.5. The degeneracy between the luminosity distance (d_L) and inclination angle (ι) parameters is resolved with an additional detector (here A1), even when the contribution of A1 in increasing the network SNR is not appreciable relative to the two detector network (L1V1). Note that, here, GW190425-like events are generated with a waveform model (IMRPhenomD), which does not include higher-order modes. Also, both the compact objects (in this case: BNS) are of approximately equal masses i.e. $q \approx 1$. Hence, we can safely assume that the higher-order modes do not play a role in the resolution of degeneracy between the parameters. We obtain similar results on relaxing the condition over \mathcal{M} and performing a similar analysis for 60 different noise realizations. An investigation addressing the luminosity distance and inclination angle degeneracy for BNS systems has also been done in [242]. It is not clear

that a better measurement of both the polarizations (h_+ & h_\times) in a larger network leads to a more precise measurement of the inclination - especially for face-on systems ($\iota < 45$ deg.). In our study, we show the result as an empirical observation for a GW190425-like event. An extensive study constraining the inclination angle with a network of GW detectors has been performed by Usman et al. [256].

The improvement in the measurements of luminosity distance has direct implications in cosmology, as mentioned in Section 5.1. The accurate measurements of inclination angle may lead to improvements in the constraints on the models for gamma-ray bursts and X-ray emissions from BNS mergers [257]. Similar improvements in the measurements of luminosity distance and inclination angles for binary black hole mergers by a three-detector network relative to a two-detector network have been obtained in [236].

5.6 Conclusion

The addition of A1 to the GW network is observed to improve the overall localization capabilities of the global detector network, even when A1 is in its early commissioning stages. To estimate the source parameters, we performed a full Bayesian PE from a lower cut-off frequency $f_{\text{low}} = 10$ Hz, which is representative of the future LVK Collaboration analysis of GW sources. We find that addition of A1 detector (at **aLIGO** O4 sensitivity) to the GW network leads to a reduction of the median $\Delta\Omega_{90\%}$ area to 5.6, 4.3, and 3.5 sq. deg. for cases where A1 is operating at 20%, 50%, and 80% duty cycles respectively, in comparison to the median $\Delta\Omega_{90\%}$ area of 6.6 sq. deg. obtained with the four detector L1H1V1K1 network for BNS sources with potential for multi-messenger follow-ups.

Our results suggest that an expanded GW detector with at an early phase A1 operating at a 20% duty cycle and operating at a weaker sensitivity (**aLIGO** O4) as compared to the other LIGO detectors (**aLIGO** A+ Design Sensitivity) is capable of localizing 64% of these BNS sources under 10 sq. deg in comparison to 56% by the four detector network. With the imminent improvement in the duty cycle and noise PSD of the A1 detector, an apparent enhancement in the localization capabilities of the GW network is observed (Refer Table 5.2). With the addition of an A1 detector to the GW network, the observation probability for the sub-networks of $k \geq 3$ detectors increases, leading to a decrease in localization uncertainties in the sky area. This allows for an optimized “tiled mode” search for post-merger emissions by telescopes such as the GROWTH India facility with a field of view of the order of ~ 0.4 sq. deg. in sky area. We show that improvements in duty cycles and noise sensitivity for A1 detector play a crucial role in enhancing the localization capabilities of the GW network. Hence, in order to get the maximal payoff from the addition of the A1 detector, efforts should be made towards maximizing the operational duty cycle and improving the noise sensitivity as soon as the detector becomes operational.

Furthermore, we show that even for BNS sources that are sub-threshold in A1, the sky-localization uncertainties with the five detector L1H1V1K1A1 network are reduced in comparison to that obtained from the four detector L1H1V1K1 network. Thus, even in a situation where A1 does not detect the BNS event independently, it plays a crucial role in pinpointing the sources

that enable a fast and efficient electromagnetic follow-up by ground and space-based telescopes.

Taking the examples of two BNS events and one NSBH event from GWTC, we show the possible source localization improvements with A1 as an additional detector in the network with real noise. For this exercise, the real noisy strain data from Virgo is used as surrogate noise in A1 detector - to simulate a scenario where the Indian detector is observing the event but has not achieved its design sensitivity.

We reaffirm the role of an additional detector (A1 in our case) in resolving the degeneracy between luminosity distance and inclination angle parameters relative to a two-detector network for a GW190425-like BNS source. This is shown by reconstructing the source parameters for GW190425-like BNS events in real, non-Gaussian noise, with the L1V1 and L1V1A1 detector networks, respectively, where data samples from V1 are used as surrogates for A1.

5.7 Discussion

In order to maximize the incentives from the GW detection of BNS sources, the EM follow-up of these events is of utmost importance. A1, joining the network of terrestrial GW detectors in the early 2030s, will enhance the localization capabilities of the network. We studied the impact of the addition of A1 in the detector network in the localization of BNS sources with moderately high signal-to-noise ratios.

The observation of an event with three or more detectors working in conjunction is fundamental for achieving localization uncertainties small enough so as to allocate telescope time for subsequent electromagnetic follow-ups. Our results presented in Fig. 5.2 from Section 5.3 can be considered optimistic, owing to the assumption of a BNS event being observed by more than two detectors at any given time. Including sub-networks of two detectors will lead to the broadening of the distribution of localization uncertainties, causing a slight shift to the right in the Cumulative Distribution Functions (CDFs) shown in Fig. 5.2. However, this is beyond the scope of this work, and a more realistic study taking the case of two detector subnetworks into account can be performed in the future. Along with this, considering the case where one or more detectors turn out to be at duty cycles that are lower than expected, as is the case for Virgo and KAGRA during the O4 run, can provide a more realistic account depicting the localization capabilities of a GW network. For instance, in the context of our study, the median $\Delta\Omega_{90\%}$ area of ~ 13.5 sq. deg. is obtained with the four detector L1H1V1K1 network, where L1 and H1 are at 80% duty cycle and both V1 and K1 detectors are operating at a lower 20% duty cycle. With the addition of A1 detector to this network, where A1 is set to aLIGO O4 sensitivity and operates at 20% duty cycle (same as that of V1 and K1), there is a significant reduction in median $\Delta\Omega_{90\%}$ area to ~ 8 sq. deg. A case study of the localization capabilities considering only the three LIGO detectors (L1, H1, and A1) is presented in [236], where all three are considered to be at A+ sensitivity.

For this study, we generate the BNS events in Section 5.1.3 using the `IMRPhenomD` waveform model and reconstruct the source parameters using the `TaylorF2` model template waveforms. Using a waveform model that includes tidal deformability parameters, higher-order modes, and other physical effects captured by additional intrinsic parameters in the analysis can make the

study more comprehensive. We aim to incorporate the tidal parameters and higher-order modes within the meshfree framework in the future. This extension will enable us to achieve a more comprehensive and rigorous analysis. We have also fixed the values of ι and ψ as shown in Section 5.1.3. This may also have an effect on the localization results. For a more general treatment, the events under consideration should be generated such that all the parameters should be allowed to vary in parameter space. This shall allow for a more exhaustive assessment of the localization capabilities of different GW networks. We simulate uncorrelated Gaussian noise in the detectors for our analysis. In this context, it has been shown by Berry et al. [258] that no appreciable impact is observed in the localization results for the case of simulated signal injected in real detector noise.

Another aspect that might affect the sky localization area evaluated from the Bayesian posterior samples is the narrow prior boundaries taken over the intrinsic parameters. For consistency, we evaluate the sky localization areas from the posterior samples with wide boundaries over intrinsic parameters using another rapid PE method (relative binning in this case) and compared the results with the meshfree framework adopted here. We find that the difference between the localization areas obtained from these two approaches is not significant enough to affect the localization results, at least for a network involving three or more detectors. The results showcased in this work serve as a demonstration of what can be accomplished by adding A1 as a new detector to the gravitational wave (GW) network. The primary emphasis is on evaluating the GW network's ability to pinpoint the source of gravitational waves, particularly in the context of potential electromagnetic follow-up observations.

Acknowledgements

We would like to thank Varun Bhalerao, Gaurav Waratkar, Aditya Vijaykumar, Sanjit Mitra, and Abhishek Sharma for useful suggestions and comments. We especially thank the anonymous referee for their careful review and helpful suggestions.

S. S. is supported by IIT Gandhinagar. L. P. is supported by the Research Scholarship Program of Tata Consultancy Services (TCS). A. S. gratefully acknowledges the generous grant provided by the Department of Science and Technology, India, through the DST-ICPS cluster project funding. We thank the HPC support staff at IIT Gandhinagar for their help and cooperation.

5.8 Appendix

5.8.1 GWTC-like events corner plots

Here are the corner plots of the three GWTC-like events, which are injected into the real noise representing O2 sensitivities. We want to study how the addition of LIGO-Aundha (A1) to the GW network observing a given GW-like event might have impacted their sky localization.

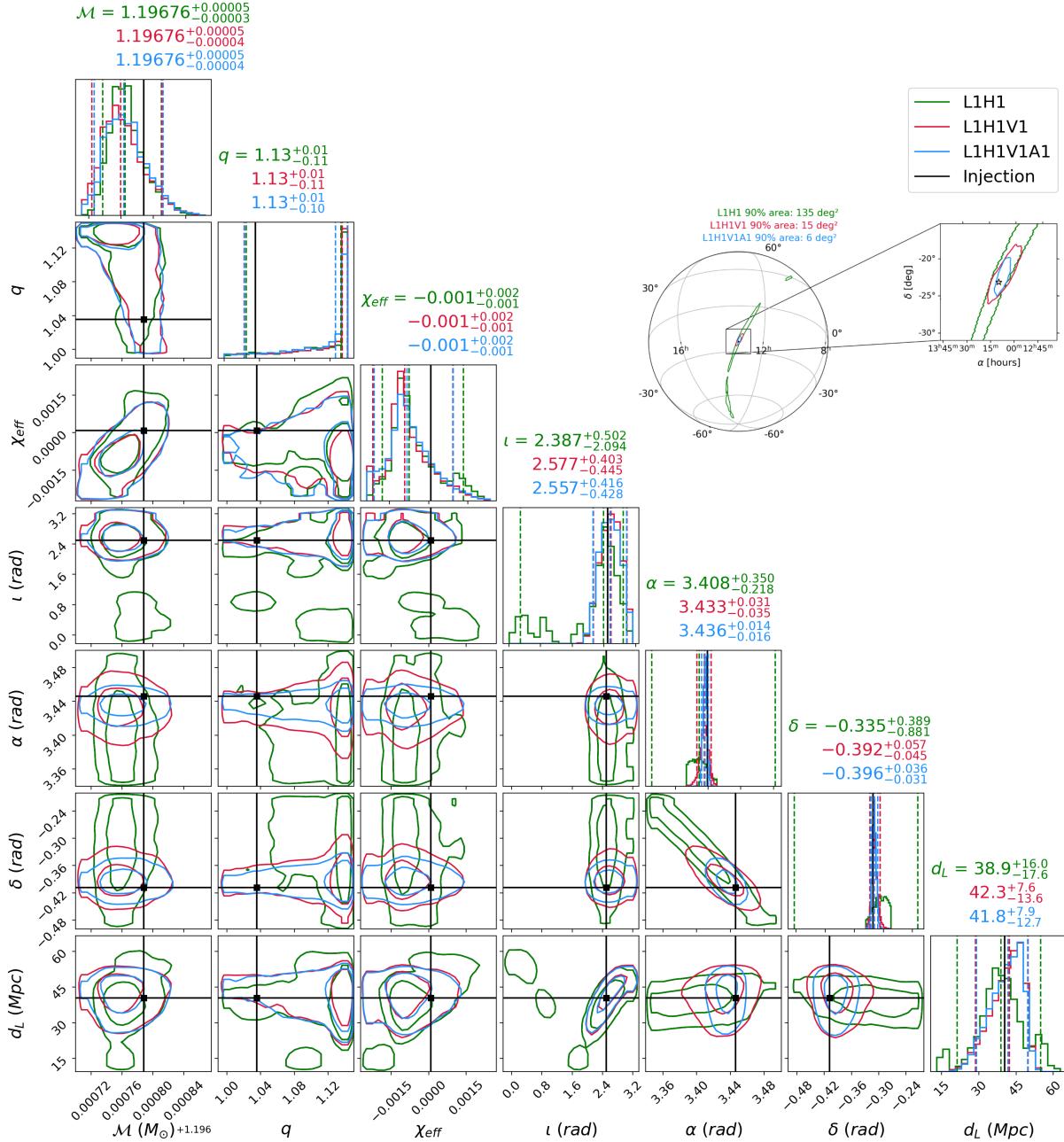


Figure 5.6: Corner plot showing the posterior distributions (with 90% CI) of various binary parameters of GW170817-like event. The black vertical line represents the corresponding injection parameters. As evident from the sky-map (in *inset*), the addition of A1 decreases the 90% credible sky localization area by $\sim 60\%$ (in comparison to L1H1V1 network) while the estimation of other parameters broadly remains similar.

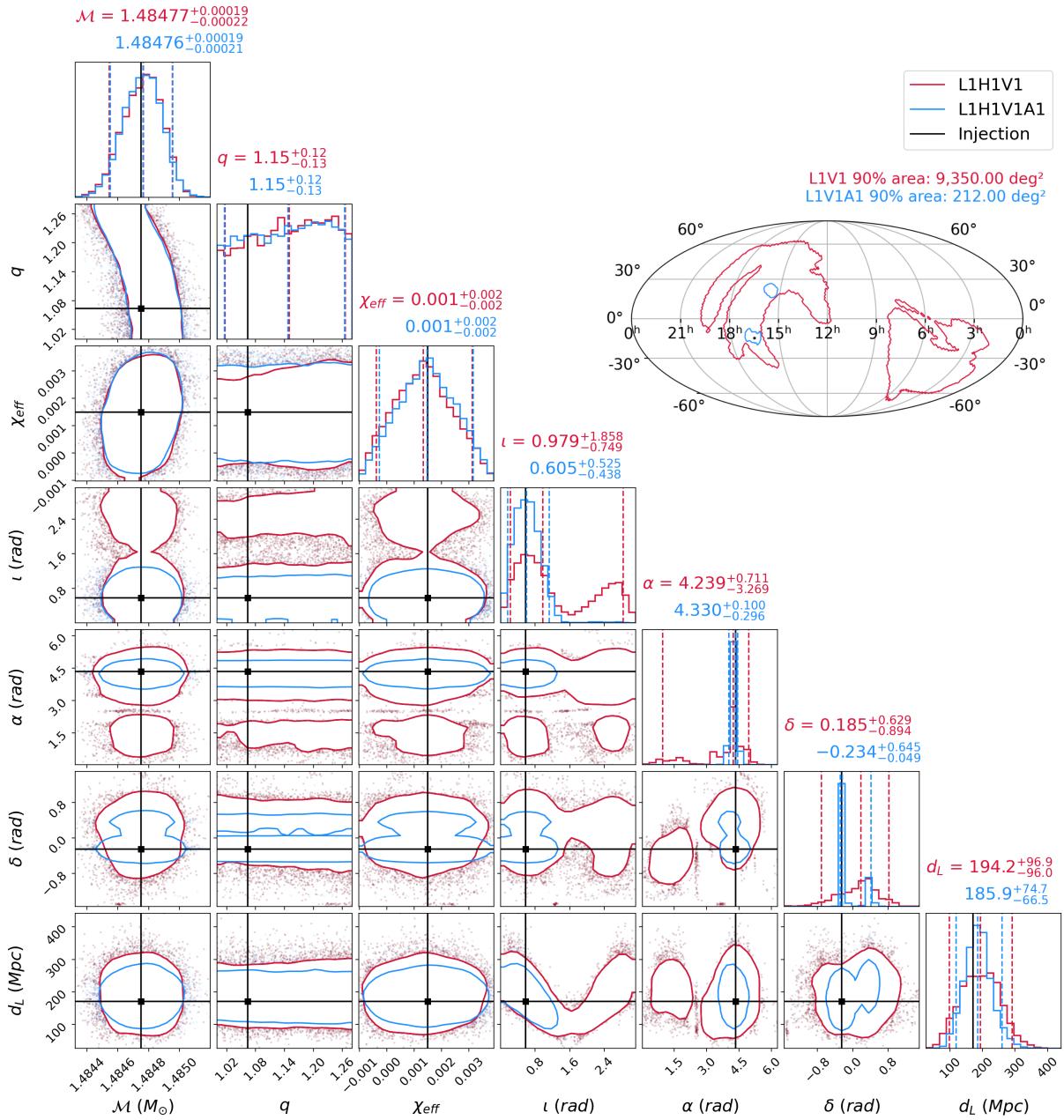


Figure 5.7: Corner plot showing the posterior distributions (with 90% CI) of various binary parameters of GW190425-like event, injected in real noise representing O2 sensitivity. The black line represents the corresponding injection parameters. As evident from the sky-map (in *inset*), the addition of A1 significantly decreases the 90% credible sky localization area by $\sim 98\%$. In addition, the degeneracy between distance and inclination is also broken, which helps constrain the distance and inclination effectively. Similar to the GW170817-like case, the estimation of other parameters broadly agrees across the two cases.

5.8.2 Simulated BNS event

Here, we present a corner plot representing the posterior distributions of various binary parameters corresponding to one of the 500 simulated BNS events injected in the L1, H1, V1, and K1 network

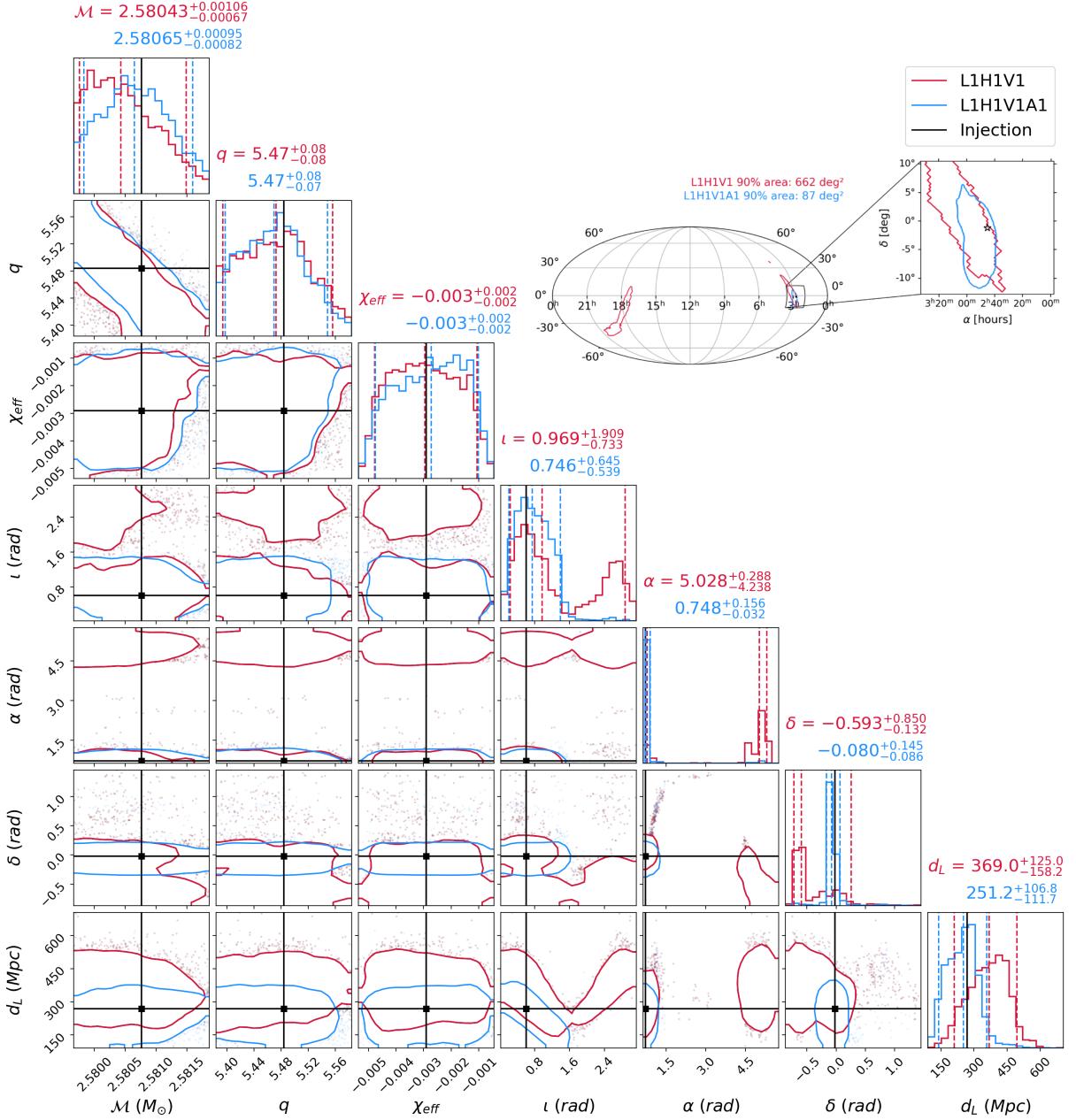


Figure 5.8: Corner plot showing the posterior distributions (with 90% CI) of various binary parameters of GW200115-like event, injected in real noise representing O2 sensitivity. The black line represents the corresponding injection parameters. As evident from the sky map (in *inset*), the addition of A1 significantly decreases the 90% credible sky localization area by $\sim 86\%$. In addition, the degeneracy between distance and inclination is again broken, which helps constrain the distance and inclination effectively. However, the estimation of other parameters broadly agrees across the two cases.

operating at O5 design sensitivity. In the same plot, we also show the posterior distributions of the binary parameters for the cases when the same event is also injected in LIGO-Aundha (A1) in addition to L1, H1, V1, and K1 detectors (at O5 sensitivities), operating at O4 and O5 design sensitivity.

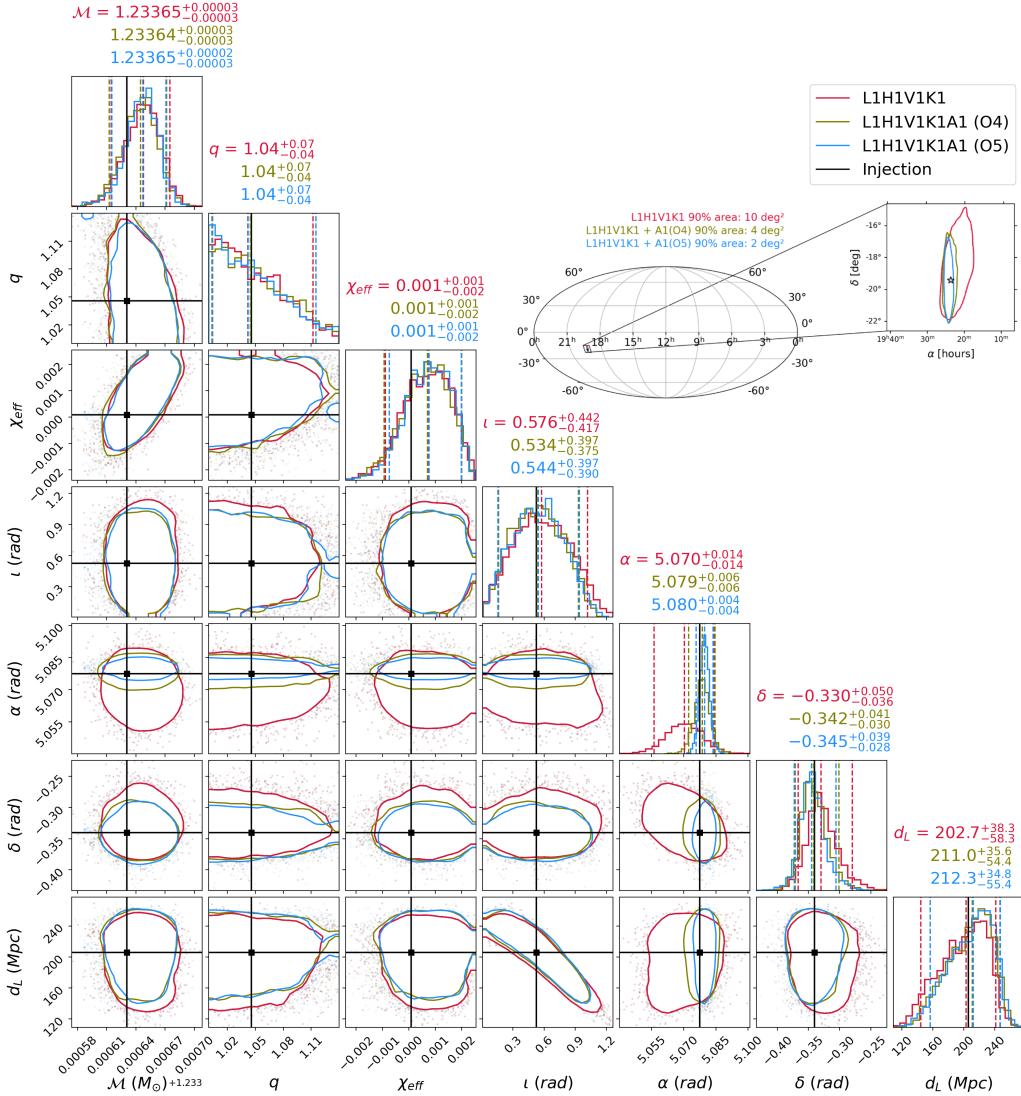


Figure 5.9: Corner plot showing the posterior distributions (with 90% CI) of various parameters of one of 500 simulations for three different cases. *Red*: When the network consists of four GW detectors, all operating at O5 design sensitivity. *Olive*: When the network consists of five GW detectors, out of which four detectors are operating at O5 design sensitivity while A1 is at O4 sensitivity. *Blue*: When the network consists of five GW detectors, all operating at O5 design sensitivity. The black vertical line represents the injection parameter. The *inset* shows the corresponding skymaps. Evidently, the 90% CI decreases from 10 sq. deg. to 4 sq. deg.(2 sq. deg.) as A1 operating at O4(O5) sensitivity is added to the L1, H1, V1, K1 network.

Chapter 6

Conclusion and Future Outlook

In this thesis, we explored rapid parameter estimation methods by utilizing a blend of singular value decomposition (SVD), a widely recognized linear algebra approach, and meshfree approximation, commonly applied in the interpolation of scattered data. The chapter 1 provides a foundational overview of Gravitational waves theory, along with a brief exploration of methods for their detection. Additionally, we delve into diverse sampling techniques commonly employed for probability distribution sampling, specifically Markov Chain Monte Carlo (MCMC) and Nested Sampling. These methodologies serve as the foundational elements for various estimation approaches utilized in the analysis of Gravitational Wave data.

The second chapter introduces a rapid method for generating waveforms that leverages SVD and meshfree approximation. This technique involves creating meshfree surrogates for GW waveform models by producing interpolations for the amplitude and phase of the waveform separately. A greedy algorithm is employed to identify an appropriate basis (for both amplitude and phase) to cover the given sample space of binary parameters (such as component masses and spins). Once these interpolations are prepared, they enable the swift generation of otherwise computationally expensive waveforms without sacrificing accuracy. To validate the reliability of our mesfhree surrogates, we apply them in reconstructing a simulated BBH source.

In chapter 3, we present a rapid parameter estimation technique that circumvents waveform generation entirely and directly interpolates the likelihood. This method consists of two stages: In the initial stage (referred to as the "start-up stage"), we employ SVD to identify a suitable basis for the space spanned by the overlap time series (between the GW strain data and templates). Subsequently, the corresponding SVD coefficients are fitted with a combination of radial basis functions (RBFs) and monomials to generate their interpolating functions. Once these functions are prepared, they are utilized to quickly evaluate the likelihood of the samples proposed by the sampler (i.e., MCMC or Nested Sampling) during the sampling process, known as the "online stage." We validated the effectiveness of our approach using a simulated BNS injected into Gaussian noise from a Power Spectral Density (PSD), emulating the design sensitivities of the LIGO-Livingston detector "L1." Our results demonstrated a substantial speed-up of over three orders of magnitude in likelihood evaluation compared to the bruteforce likelihood calculation implemented in PYCBC [15, 16] while showing good agreement between the meshfree and bruteforce likelihoods.

In chapter 4, we expand upon the meshfree framework introduced in chapter 3 to enable coherent network parameter estimation (PE) by incorporating gravitational wave (GW) strain data from multiple detectors. This extension facilitates the estimation of various parameters, including sky location and inclination, crucial for follow-ups of the electromagnetic counterparts of such events. Additionally, we propose an effective strategy for node placement to distribute nodes in the intrinsic parameter space. This placement is guided by the Fisher matrix calculated at the center of the selected sample space box, enhancing node concentration near the likelihood peak for higher accuracy in meshfree likelihood approximation. To validate our method’s robustness, we inject numerous simulated binary neutron star (BNS) signals into Gaussian noise, mimicking the detector sensitivities akin to the theoretical design sensitivities of ground-based detectors such as LIGO-Virgo. We estimate their source parameters using our meshfree approach and observe substantial agreement with the relative binning method [139, 138], an alternative rapid PE method employed in the GW community. Considering that BNS systems are influenced by tidal effects, we conduct a preliminary analysis within a meshfree framework, incorporating tidal parameters for both components of a BNS system. This inclusion increases the interpolating parameter space dimensionality to six, necessitating a higher number of nodes for an accurate likelihood approximation. We again observe a broad agreement between the posterior distributions of binary parameters in PE analyses conducted with both meshfree and bruteforce likelihood methods.

In chapter 5, we employ our meshfree-based rapid parameter estimation (PE) pipeline to conduct an extensive study of the potential impact of LIGO-Aundha. LIGO-Aundha is an upcoming Indian Gravitational Waves Observatory located in Hingoli, Maharashtra, expected to commence scientific operations around 2030. This observatory is poised to join the international gravitational wave detector network (IGWN). The focus of our study is on evaluating the influence of LIGO-Aundha joining the IGWN on the sky-localization capabilities of BNS systems, considering their significance as potential sources for electromagnetic (EM) follow-ups. Our findings suggest that, even with moderate sensitivity, the addition of LIGO-Aundha has the potential to significantly enhance the overall scientific objectives of second-generation gravitational wave observatories.

Finally, we highlight some of the potential extensions of the rapid PE methods presented in this thesis. In the current implementation, as outlined in chapter 2, the waveform interpolation primarily addresses aligned spin waveform models, specifically focusing on the dominant “quadrupolar” mode. In future iterations, our aim is to broaden the scope of our method to encompass precessing waveform models and models that include subdominant modes. The likelihood interpolation technique, initially introduced in chapter 3 and extended for coherent network parameter estimation in chapter 4, currently requires a narrow region within the intrinsic parameter space for accurate interpolation of SVD coefficients and template norm squares. Additionally, it focuses on aligned spin waveform models featuring only the dominant mode. In the subsequent development of these works, we intend to overcome these limitations by proposing robust strategies for node placement, expanding boundaries in the intrinsic parameter space, and accommodating waveform models that incorporate both spin precession and sub-dominant modes.

Some of the ongoing efforts include the incorporation of eccentricity into our meshfree framework for both waveform and likelihood interpolation. Furthermore, the techniques introduced in this thesis hold promise for application to parameter estimation challenges in scenarios involving overlapping signals [259, 260, 261] in third-generation detectors like Cosmic Explorer and the Einstein Telescope. Additionally, they could find utility in addressing the parameter estimation challenges in the context of the Laser Interferometer Space Antenna (LISA) where the signal duration is $\mathcal{O}(1\text{ year})$ and a rapid PE method becomes essential for timely estimation of the source parameters.

Bibliography

- [1] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and others. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 116:061102, Feb 2016.
- [2] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al. Properties of the binary black hole merger gw150914. *Phys. Rev. Lett.*, 116:241102, Jun 2016.
- [3] Albert Einstein. The foundation of the general theory of relativity. *Annalen Phys.*, 49(7):769–822, 1916.
- [4] Albert Einstein. Approximative Integration of the Field Equations of Gravitation. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, 1916:688–696, 1916.
- [5] Albert Einstein. Über Gravitationswellen. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, 1918:154–167, 1918.
- [6] J.B. Hartle. *Gravity: An Introduction to Einstein’s General Relativity*. Cambridge University Press, 2021.
- [7] M. Maggiore. *Gravitational Waves: Volume 1: Theory and Experiments*. Gravitational Waves. OUP Oxford, 2008.
- [8] John T. Whelan. Lectures on gravitational wave data analysis, school of gravitational waves, warsaw, July 2013.
- [9] Pau Amaro-Seoane, Heather Audley, Stanislav Babak, John Baker, Enrico Barausse, Peter Bender, Emanuele Berti, Pierre Binetruy, Michael Born, Daniele Bortoluzzi, et al. Laser interferometer space antenna, 2017.
- [10] Travis Robson, Neil J Cornish, and Chang Liu. The construction and use of lisa sensitivity curves. *Classical and Quantum Gravity*, 36(10):105011, April 2019.
- [11] Neil J. Cornish and Kevin Shuman. Black hole hunting with lisa. *Physical Review D*, 101(12), jun 2020.
- [12] J Aasi, B P Abbott, R Abbott, T Abbott, M R Abernathy, K Ackley, C Adams, T Adams, P Addesso, R X Adhikari, et al. Advanced ligo. *Classical and Quantum Gravity*, 32(7):074001, March 2015.

- [13] Benjamin P Abbott, Rich Abbott, TD Abbott, Fausto Acernese, Kendall Ackley, Carl Adams, Thomas Adams, Paolo Addesso, RX Adhikari, Vaishali B Adya, et al. GW170817: observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett.*, 119(16):161101, 2017.
- [14] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, et al. Properties of the binary neutron star merger gw170817. *Phys. Rev. X*, 9:011001, Jan 2019.
- [15] Samantha A Usman, Alexander H Nitz, Ian W Harry, Christopher M Biwer, Duncan A Brown, Miriam Cabero, Collin D Capano, Tito Dal Canton, Thomas Dent, Stephen Fairhurst, et al. The PyCBC search for gravitational waves from compact binary coalescence. *Class. Quantum Gravity*, 33(21):215004, 2016.
- [16] Christopher Michael Biwer, Collin D Capano, Soumi De, Miriam Cabero, Duncan A Brown, Alexander H Nitz, and Vivien Raymond. PyCBC Inference: a Python-based parameter estimation toolkit for compact binary coalescence signals. *Publ. Astron. Soc. Pac.*, 131(996):024503, 2019.
- [17] Luc Blanchet, Thibault Damour, and Bala R. Iyer. Gravitational waves from inspiralling compact binaries: Energy loss and waveform to second-post-newtonian order. *Phys. Rev. D*, 51:5360–5386, May 1995.
- [18] Luc Blanchet, Guillaume Faye, Bala R Iyer, and Siddhartha Sinha. The third post-newtonian gravitational wave polarizations and associated spherical harmonic modes for inspiralling compact binaries in quasi-circular orbits. *Classical and Quantum Gravity*, 25(16):165003, jul 2008.
- [19] K.G. Arun, Luc Blanchet, Bala.R. Iyer, and Siddhartha Sinha. Third post-newtonian angular momentum flux and the secular evolution of orbital elements for inspiralling compact binaries in quasi-elliptical orbits. *Physical Review D*, 80(12), dec 2009.
- [20] Luc Blanchet. Gravitational radiation from post-newtonian sources and inspiralling compact binaries. *Living Reviews in Relativity*, 17(1), February 2014.
- [21] R. A. Hulse and J. H. Taylor. Discovery of a pulsar in a binary system. *The Astrophysical Journal Letters*, 195:L51–L53, Jan 1975.
- [22] Russell A. Hulse. The discovery of the binary pulsar. *Rev. Mod. Phys.*, 66:699–710, Jul 1994.
- [23] Joseph H. Taylor. Binary pulsars and relativistic gravity. *Rev. Mod. Phys.*, 66:711–719, Jul 1994.
- [24] I.A. Gerasimov. Book Review: N. T. Roseveare Mercury’s Perihelion from Le Verrier to Einstein. *sovast*, 30:365, jun 1986.

- [25] Clifford M. Will. New general relativistic contribution to mercury's perihelion advance. *Physical Review Letters*, 120(19), May 2018.
- [26] Irwin I. Shapiro. Fourth test of general relativity. *Phys. Rev. Lett.*, 13:789–791, Dec 1964.
- [27] Michael Kramer. Determination of the geometry of the psr b1913+16 system by geodetic precession. *The Astrophysical Journal*, 509(2):856, dec 1998.
- [28] Joel M. Weisberg and Joseph H. Taylor. General relativistic geodetic spin precession in binary pulsar b1913+16: Mapping the emission beam in two dimensions. *The Astrophysical Journal*, 576(2):942, sep 2002.
- [29] J. M. Weisberg and J. H. Taylor. Relativistic binary pulsar b1913+16: Thirty years of observations and analysis, 2004.
- [30] Benjamin P Abbott, Richard Abbott, TD Abbott, MR Abernathy, Fausto Acernese, Kendall Ackley, Carl Adams, Thomas Adams, Paolo Addesso, RX Adhikari, et al. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 116(6):061102, 2016.
- [31] Alessandro Manzotti and Alexander Dietz. Prospects for early localization of gravitational-wave signals from compact binary coalescences with advanced detectors, 2012.
- [32] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, et al. Multi-messenger Observations of a Binary Neutron Star Merger. *The*, 848(2):L12, October 2017.
- [33] P. A. Evans, S. B. Cenko, J. A. Kennea, S. W. K. Emery, N. P. M. Kuin, O. Korobkin, R. T. Wollaeger, C. L. Fryer, K. K. Madsen, F. A. Harrison, et al. swift and nustar observations of gw170817: Detection of a blue kilonova. *Science*, 358(6370):1565–1570, 2017.
- [34] E. Troja, L. Piro, H. van Eerten, R. T. Wollaeger, M. Im, O. D. Fox, N. R. Butler, S. B. Cenko, T. Sakamoto, C. L. Fryer, et al. The x-ray counterpart to the gravitational-wave event gw170817. *Nature*, 551(7678):71–74, 2017.
- [35] Bruce Allen, Warren G. Anderson, Patrick R. Brady, Duncan A. Brown, and Jolien D. E. Creighton. Findchirp: An algorithm for detection of gravitational waves from inspiraling compact binaries. *Physical Review D*, 85(12), jun 2012.
- [36] G. Turin. An introduction to matched filters. *IRE Transactions on Information Theory*, 6(3):311–329, 1960.
- [37] A. A. Michelson. The relative motion of the Earth and of the luminiferous ether. *American Journal of Science*, 22(128):120–129, aug 1881.
- [38] A. A. Michelson and E. W. Morley. On the relative motion of the Earth and the luminiferous ether. *American Journal of Science*, 34(203):333–345, nov 1887.

- [39] Ch. Fabry and A. Perot. On a New Form of Interferometer. *The Astrophysical Journal*, 13:265, may 1901.
- [40] A. Perot and Charles Fabry. On the Application of Interference Phenomena to the Solution of Various Problems of Spectroscopy and Metrology. *The Astrophysical Journal*, 9:87, feb 1899.
- [41] J Aasi, B P Abbott, R Abbott, T Abbott, M R Abernathy, K Ackley, C Adams, T Adams, P Addesso, R X Adhikari, et al. Advanced LIGO. *Classical and Quantum Gravity*, 32(7):074001, mar 2015.
- [42] B P Abbott, R Abbott, T D Abbott, S Abraham, F Acernese, K Ackley, C Adams, V B Adya, C Affeldt, M Agathos, et al. A guide to ligo–virgo detector noise and extraction of transient gravitational-wave signals. *Classical and Quantum Gravity*, 37(5):055002, February 2020.
- [43] Kentaro Somiya. Detector configuration of KAGRA—the Japanese cryogenic gravitational-wave detector. *Classical and Quantum Gravity*, 29(12):124007, June 2012.
- [44] Yoichi Aso, Yuta Michimura, Kentaro Somiya, Masaki Ando, Osamu Miyakawa, Takanori Sekiguchi, Daisuke Tatsumi, and Hiroaki Yamamoto. Interferometer design of the KAGRA gravitational wave detector. *Phys. Rev. D*, 88:043007, Aug 2013.
- [45] T Akutsu, M Ando, K Arai, Y Arai, S Araki, A Araya, N Aritomi, Y Aso, S Bae, Y Bae, et al. Overview of KAGRA: Detector design and construction history. *Progress of Theoretical and Experimental Physics*, 2021(5):05A101, 08 2020.
- [46] Benjamin P Abbott et al. Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Rev. Relativ.*, 23(1):1–69, 2020.
- [47] Bala Iyer, Tarun Souradeep, CS Unnikrishnan, Sanjeev Dhurandhar, Sendhil Raja, and Anand Sengupta. LIGO India - Proposal of the Consortium for Indian Initiative in Gravitational-wave Observations (IndIGO). Technical Report M1100296-v2, LIGO Scientific Collaboration, 2011.
- [48] Shivaraj Kandhasamy and Sukanta Bose. LIGO-India observatory coordinate system for GW analyses. Technical Report T2000158-v3, LIGO Scientific Collaboration, 2023.
- [49] C. S. Unnikrishnan. IndIGO and LIGO-India: scope and plans for gravitational wave research and precision metrology in india. *International Journal of Modern Physics D*, 22(01):1341010, jan 2013.
- [50] C. S. Unnikrishnan. LIGO-India: A Decadal Assessment on Its Scope, Relevance, Progress, and Future, 2023.

- [51] M. Punturo, M. Abernathy, F. Acernese, B. Allen, N. Andersson, K. Arun, F. Barone, B. Barr, M. Barsuglia, M. Beker, et al. The third generation of gravitational wave observatories and their science reach. *Classical and Quantum Gravity*, 27(8):084007, apr 2010.
- [52] S. Hild, M. Abernathy, F. Acernese, P. Amaro-Seoane, N. Andersson, K. Arun, F. Barone, B. Barr, M. Barsuglia, M. Beker, et al. Sensitivity studies for third-generation gravitational wave observatories. *Classical and Quantum Gravity*, 28(9):094013, may 2011.
- [53] B P Abbott, R Abbott, T D Abbott, M R Abernathy, K Ackley, C Adams, P Addesso, R X Adhikari, V B Adya, C Affeldt, et al. Exploring the sensitivity of next generation gravitational wave detectors. *Classical and Quantum Gravity*, 34(4):044001, jan 2017.
- [54] David Reitze, Rana X Adhikari, Stefan Ballmer, Barry Barish, Lisa Barsotti, GariLynn Billingsley, Duncan A. Brown, Yanbei Chen, Dennis Coyne, Robert Eisenstein, et al. Cosmic explorer: The u.s. contribution to gravitational-wave astronomy beyond ligo, 2019.
- [55] Stanislav Babak, Martin Hewitson, and Antoine Petiteau. Lisa sensitivity and snr calculations, 2021.
- [56] Seiji Kawamura, Masaki Ando, Naoki Seto, Shuichi Sato, Takashi Nakamura, Kimio Tsubono, Nobuyuki Kanda, Takahiro Tanaka, Jun'ichi Yokoyama, Ikkoh Funaki, et al. The Japanese space gravitational wave antenna: DECIGO. *Classical and Quantum Gravity*, 28(9):094011, may 2011.
- [57] Seiji Kawamura, Masaki Ando, Naoki Seto, Shuichi Sato, Mitsuru Musha, Isao Kawano, Jun'ichi Yokoyama, Takahiro Tanaka, Kunihito Ioka, Tomotada Akutsu, et al. Current status of space gravitational wave antenna decigo and b-decigo, 2020.
- [58] Shuichi Sato, Seiji Kawamura, Masaki Ando, Takashi Nakamura, Kimio Tsubono, Akito Araya, Ikkoh Funaki, Kunihito Ioka, Nobuyuki Kanda, Shigenori Moriwaki, et al. The status of decigo. *Journal of Physics: Conference Series*, 840(1):012010, may 2017.
- [59] C J Moore, R H Cole, and C P L Berry. Gravitational-wave sensitivity curves. *Classical and Quantum Gravity*, 32(1):015014, December 2014.
- [60] G. Casella and R.L. Berger. *Statistical Inference*. Duxbury advanced series in statistics and decision sciences. Thomson Learning, 2002.
- [61] Martin T. Wells George Casella, Christian P. Robert. Generalized accept-reject sampling schemes. *Lecture Notes-Monograph Series*, 45:342–347, 2004.
- [62] T. Kloek and H. K. van Dijk. Bayesian estimates of equation system parameters: An application of integration by monte carlo. *Econometrica*, 46(1):1–19, 1978.
- [63] William Feller. *An introduction to probability theory and its applications, Volume 2*, volume 81. John Wiley & Sons, 1991.

- [64] George Casella and Edward I George. Explaining the gibbs sampler. *The American Statistician*, 46(3):167–174, 1992.
- [65] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013.
- [66] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21(6):1087–1092, jun 1953.
- [67] Marshall N. Rosenbluth. Genesis of the Monte Carlo Algorithm for Statistical Mechanics. In James E. Gubernatis, editor, *The Monte Carlo Method in the Physical Sciences*, volume 690 of *American Institute of Physics Conference Series*, pages 22–30, Nov 2003.
- [68] Bernd A. Berg. Introduction to markov chain monte carlo simulations and their statistical analysis, 2004.
- [69] W.R. Gilks, S. Richardson, and D. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Chapman & Hall/CRC Interdisciplinary Statistics. Taylor & Francis, 1995.
- [70] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2005.
- [71] John Skilling. Nested sampling for general Bayesian computation. *Bayesian Anal.*, 1(4):833–859, 2006.
- [72] Joshua S Speagle. DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Mon. Not. R. Astron. Soc.*, 493(3):3132–3158, 2020.
- [73] Sergey Koposov, Josh Speagle, Kyle Barbary, Gregory Ashton, Ed Bennett, Johannes Buchner, Carl Scheffler, Ben Cook, Colm Talbot, James Guillochon, et al. [joshspeagle/dynesty: v2.1.0](#), Feb 2023.
- [74] Edward Higson, Will Handley, Michael Hobson, and Anthony Lasenby. Dynamic nested sampling: an improved algorithm for parameter estimation and evidence calculation. *Statistics and Computing*, 29(5):891–913, 2019.
- [75] D. Sivia and J. Skilling. *Data Analysis: A Bayesian Tutorial*. Oxford science publications. OUP Oxford, 2006.
- [76] R. McElreath. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. A Chapman & Hall book. CRC Press, 2020.
- [77] Sebastian Khan, Sascha Husa, Mark Hannam, Frank Ohme, Michael Pürrer, Xisco Jiménez Forteza, and Alejandro Bohé. Frequency-domain gravitational waves from nonprecessing black-hole binaries. ii. a phenomenological model for the advanced detector era. *Physical Review D*, 93(4), February 2016.

- [78] Geraint Pratten, Sascha Husa, Cecilio García-Quirós, Marta Colleoni, Antoni Ramos-Buades, Héctor Estellés, and Rafel Jaume. Setting the cornerstone for a family of models for gravitational waves from compact binaries: The dominant harmonic for nonprecessing quasicircular black holes. *Physical Review D*, 102(6), September 2020.
- [79] Sebastian Khan, Katerina Chatzioannou, Mark Hannam, and Frank Ohme. Phenomenological model for the gravitational-wave signal from precessing binary black holes with two-spin effects. *Physical Review D*, 100(2), July 2019.
- [80] Cecilio García-Quirós, Marta Colleoni, Sascha Husa, Héctor Estellés, Geraint Pratten, Antoni Ramos-Buades, Maite Mateu-Lucena, and Rafel Jaume. Multimode frequency-domain model for the gravitational wave signal from nonprecessing black-hole binaries. *Physical Review D*, 102(6), September 2020.
- [81] Geraint Pratten, Cecilio Garcí a-Quirós, Marta Colleoni, Antoni Ramos-Buades, Héctor Estellés, Maite Mateu-Lucena, Rafel Jaume, Maria Haney, David Keitel, Jonathan E. Thompson, et al. Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes. *Physical Review D*, 103(10), may 2021.
- [82] Alejandro Bohé, Lijing Shao, Andrea Taracchini, Alessandra Buonanno, Stanislav Babak, Ian W. Harry, Ian Hinder, Serguei Ossokine, Michael Pürrer, Vivien Raymond, Tony Chu, Heather Fong, Prayush Kumar, Harald P. Pfeiffer, Michael Boyle, Daniel A. Hemberger, Lawrence E. Kidder, Geoffrey Lovelace, Mark A. Scheel, and Béla Szilágyi. Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors. *Physical Review D*, 95(4), February 2017.
- [83] Roberto Cotesta, Sylvain Marsat, and Michael Pürrer. Frequency-domain reduced-order model of aligned-spin effective-one-body waveforms with higher-order modes. *Physical Review D*, 101(12), June 2020.
- [84] Jonathan Blackman, Scott E. Field, Mark A. Scheel, Chad R. Galley, Christian D. Ott, Michael Boyle, Lawrence E. Kidder, Harald P. Pfeiffer, and Béla Szilágyi. Numerical relativity waveform surrogate model for generically precessing binary black hole mergers. *Physical Review D*, 96(2), July 2017.
- [85] Vijay Varma, Scott E. Field, Mark A. Scheel, Jonathan Blackman, Davide Gerosa, Leo C. Stein, Lawrence E. Kidder, and Harald P. Pfeiffer. Surrogate models for precessing binary black hole simulations with unequal masses. *Physical Review Research*, 1(3), October 2019.
- [86] Vijay Varma, Scott E. Field, Mark A. Scheel, Jonathan Blackman, Lawrence E. Kidder, and Harald P. Pfeiffer. Surrogate model of hybridized numerical relativity binary black hole waveforms. *Physical Review D*, 99(6), March 2019.
- [87] Kipp Cannon, Chad Hanna, and Drew Keppel. Interpolating compact binary waveforms using the singular value decomposition. *Phys. Rev. D*, 85(8):081504, 2012.

- [88] P. Ajith, M. Hannam, S. Husa, Y. Chen, B. Brügmann, N. Dorband, D. Müller, F. Ohme, D. Pollney, C. Reisswig, L. Santamaría, and J. Seiler. Inspiral-merger-ringdown waveforms for black-hole binaries with nonprecessing spins. *Physical Review Letters*, 106(24), jun 2011.
- [89] Alvin J. K. Chua, Chad R. Galley, and Michele Vallisneri. Reduced-order modeling with artificial neurons for gravitational-wave inference. *Phys. Rev. Lett.*, 122:211101, May 2019.
- [90] K. G. Arun, Alessandra Buonanno, Guillaume Faye, and Evan Ochsner. Higher-order spin effects in the amplitude and phase of gravitational waveforms emitted by inspiraling compact binaries: Ready-to-use gravitational waveforms. *Phys. Rev. D*, 79:104023, May 2009.
- [91] Sebastian Khan and Rhys Green. Gravitational-wave surrogate models powered by artificial neural networks. *Physical Review D*, 103(6), mar 2021.
- [92] D. Williams, I.S. Heng, J. Gair, J.A. Clark, and B. Khamesra. Precessing numerical relativity waveform surrogate model for binary black holes: A gaussian process regression approach. *Physical Review D*, 101(6), mar 2020.
- [93] Joongoo Lee, Sang Hoon Oh, Kyungmin Kim, Gihyuk Cho, John J. Oh, Edwin J. Son, and Hyung Mok Lee. Deep learning model on gravitational waveforms in merging and ringdown phases of binary black hole coalescences. *Physical Review D*, 103(12), jun 2021.
- [94] R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, N. Adhikari, R. X. Adhikari, V. B. Adya, C. Affeldt, D. Agarwal, et al. Population of merging compact binaries inferred using gravitational waves through gwtc-3. *Phys. Rev. X*, 13:011048, Mar 2023.
- [95] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya, C. Affeldt, M. Agathos, et al. Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Reviews in Relativity*, 23(1), September 2020.
- [96] E. T. Newman and R. Penrose. Note on the Bondi-Metzner-Sachs group. *J. Math. Phys.*, 7:863–870, 1966.
- [97] J. N. Goldberg, A. J. MacFarlane, E. T. Newman, F. Rohrlich, and E. C. G. Sudarshan. Spin s spherical harmonics and edth. *J. Math. Phys.*, 8:2155, 1967.
- [98] S. V. Dhurandhar and B. S. Sathyaprakash. Choice of filters for the detection of gravitational waves from coalescing binaries. ii. detection in colored noise. *Phys. Rev. D*, 49:1707–1722, Feb 1994.
- [99] Serge Droz, Daniel J. Knapp, Eric Poisson, and Benjamin J. Owen. Gravitational waves from inspiraling compact binaries: Validity of the stationary-phase approximation to the Fourier transform. *Phys. Rev. D*, 59:124016, May 1999.

- [100] Guillaume Faye, Sylvain Marsat, Luc Blanchet, and Bala R Iyer. The third and a half-post-newtonian gravitational wave quadrupole mode for quasi-circular inspiralling compact binaries. *Classical and Quantum Gravity*, 29(17):175004, aug 2012.
- [101] Luc Blanchet, Thibault Damour, Bala R. Iyer, Clifford M. Will, and Alan G. Wiseman. Gravitational-radiation damping of compact binary systems to second post-newtonian order. *Phys. Rev. Lett.*, 74:3515–3518, May 1995.
- [102] Sebastian Khan, Sascha Husa, Mark Hannam, Frank Ohme, Michael Pürer, Xisco Jiménez Forteza, and Alejandro Bohé. Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era. *Phys. Rev. D*, 93(4):044007, 2016.
- [103] Sascha Husa, Sebastian Khan, Mark Hannam, Michael Pürer, Frank Ohme, Xisco Jiménez Forteza, and Alejandro Bohé. Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal. *Phys. Rev. D*, 93(4):044006, 2016.
- [104] F.F. Kuo and J.F. Kaiser. *System Analysis by Digital Computer*. Wiley, 1966.
- [105] Michael Pürer. Frequency-domain reduced order models for gravitational waves from aligned-spin compact binaries. *Classical and Quantum Gravity*, 31(19):195010, September 2014.
- [106] Soichiro Morisaki, Rory Smith, Leo Tsukada, Surabhi Sachdev, Simon Stevenson, Colm Talbot, and Aaron Zimmerman. Rapid localization and inference on compact binary coalescences with the Advanced LIGO-Virgo-KAGRA gravitational-wave detector network, 2023.
- [107] Gregory E Fasshauer. *Meshfree Approximation Methods with Matlab*. World Scientific, 2007.
- [108] Natasha Flyer, Bengt Fornberg, Victor Bayona, and Gregory A. Barnett. On the role of polynomials in RBF-FD approximations: I. Interpolation and accuracy. *Journal of Computational Physics*, 321:21–38, September 2016.
- [109] Matthew Brand. Fast low-rank modifications of the thin singular value decomposition. *Linear Algebra and its Applications*, 415(1):20–30, 2006. Special Issue on Large Scale Linear and Nonlinear Eigenvalue Problems.
- [110] Peter Stange. On the efficient update of the singular value decomposition. *PAMM*, 8, 2008.
- [111] Priscilla Canizares, Scott E Field, Jonathan R Gair, and Manuel Tiglio. Gravitational wave parameter estimation with compressed likelihood evaluations. *Phys. Rev. D*, 87(12):124005, 2013.

- [112] Priscilla Canizares, Scott E Field, Jonathan Gair, Vivien Raymond, Rory Smith, and Manuel Tiglio. Accelerated gravitational wave parameter estimation with reduced order modeling. *Phy. Rev. Lett.*, 114(7):071104, 2015.
- [113] Soichiro Morisaki and Vivien Raymond. Rapid parameter estimation of gravitational waves from binary neutron star coalescence using focused reduced order quadrature. *Physical Review D*, 102(10), nov 2020.
- [114] Gonzalo Morras, Jose Francisco Nuno Siles, and Juan Garcia-Bellido. Efficient reduced order quadrature construction algorithms for fast gravitational wave inference, 7 2023.
- [115] L. Barsotti, S. Gras, M. Evans, and P. Fritschel. The updated Advanced LIGO design curve. Technical Report LIGO-T1800044-v5, LIGO Scientific Collaboration, 2018.
- [116] Alvin JK Chua, Chad R Galley, and Michele Vallisneri. Reduced-order modeling with artificial neurons for gravitational-wave inference. *Phys. Rev. Lett.*, 122(21):211101, 2019.
- [117] T. Hastie, R. Tibshirani, and J.H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. Springer, 2009.
- [118] Trevor Hines. Python package containing the tools necessary for radial basis function (RBF) applications, 2015.
- [119] Maximilian Dax, Stephen R. Green, Jonathan Gair, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf. Real-time gravitational wave science with neural posterior estimation. *Phys. Rev. Lett.*, 127:241103, Dec 2021.
- [120] Michael J. Williams, John Veitch, and Chris Messenger. Nested sampling with normalizing flows for gravitational-wave inference. *Phys. Rev. D*, 103:103006, May 2021.
- [121] Maxime Barrault, Yvon Maday, Ngoc Cuong Nguyen, and Anthony T. Patera. An ‘empirical interpolation’ method: application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathematique*, 339(9):667–672, 2004.
- [122] Yvon Maday, Ngoc Cuong Nguyen, Anthony T. Patera, and S. H. Pau. A general multipurpose interpolation procedure: the magic points. *Communications on Pure and Applied Analysis*, 8(1):383–404, 2009.
- [123] Saifon Chaturantabut and Danny C. Sorensen. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing*, 32(5):2737–2764, 2010.
- [124] Soichiro Morisaki. Accelerating parameter estimation of gravitational waves from compact binary coalescence using adaptive frequency resolutions. *Phys. Rev. D*, 104(4):044062, 2021.
- [125] gwosc.org, 2024.
- [126] L. Pathak. Github repository for meshfree nr surrogates, Dec 2023.

- [127] Benjamin P Abbott et al. Gravitational waves and gamma-rays from a binary neutron star merger: GW170817 and GRB 170817A. *Astrophys. J. Lett.*, 848(2):L13, 2017.
- [128] Leo P Singer and Larry R Price. Rapid Bayesian position reconstruction for gravitational-wave transients. *Phys. Rev. D*, 93(2):024013, 2016.
- [129] Daniel Finstad and Duncan A. Brown. Fast parameter estimation of binary mergers for multimessenger follow-up. *Astrophys. J. Lett.*, 905(1):L9, dec 2020.
- [130] Ben Margalit and Brian D. Metzger. The multi-messenger matrix: The future of neutron star merger constraints on the nuclear equation of state. *Astrophys. J. Lett.*, 880(1):L15, jul 2019.
- [131] Maximilian Dax, Stephen R. Green, Jonathan Gair, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf. Real-time gravitational wave science with neural posterior estimation. *Phys. Rev. Lett.*, 127:241103, Dec 2021.
- [132] Hunter Gabbard, Chris Messenger, Ik Siong Heng, Francesco Tonolini, and Roderick Murray-Smith. Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy. *Nature Physics*, 18(1):112–117, dec 2021.
- [133] Maximilian Dax, Stephen R. Green, Jonathan Gair, Michael Pürrer, Jonas Wildberger, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf. Neural importance sampling for rapid and reliable gravitational-wave inference. *Phys. Rev. Lett.*, 130:171403, Apr 2023.
- [134] Hong Qi and Vivien Raymond. Python-based reduced order quadrature building code for fast gravitational wave inference. *Phys. Rev. D*, 104(6), sep 2021.
- [135] Soichiro Morisaki and Vivien Raymond. Rapid parameter estimation of gravitational waves from binary neutron star coalescence using focused reduced order quadrature. *Phys. Rev. D*, 102:104020, Nov 2020.
- [136] Alvin JK Chua and Michele Vallisneri. Learning Bayesian posteriors with neural networks for gravitational-wave inference. *Phys. Rev. Lett.*, 124(4):041102, 2020.
- [137] Jacob Lange, Richard O’Shaughnessy, and Monica Rizzo. Rapid and accurate parameter inference for coalescing, precessing compact binaries, 2018.
- [138] Neil J. Cornish. Heterodyned likelihood for rapid gravitational wave parameter inference. *Phys. Rev. D*, 104:104054, Nov 2021.
- [139] Barak Zackay, Liang Dai, and Tejaswi Venumadhav. Relative binning and fast likelihood evaluation for gravitational wave parameter estimation, 2018.
- [140] RJE Smith, Chad Hanna, Ilya Mandel, and Alberto Vecchio. Rapidly evaluating the compact-binary likelihood function via interpolation. *Phys. Rev. D*, 90(4):044074, 2014.

- [141] Kipp Cannon, Adrian Chapman, Chad Hanna, Drew Keppel, Antony C. Searle, and Alan J. Weinstein. Singular value decomposition applied to compact binary coalescence gravitational-wave signals. *Phys. Rev. D*, 82:044025, Aug 2010.
- [142] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Fourth edition, 2013.
- [143] Daniel Foreman-Mackey, David W. Hogg, Dustin Lang, and Jonathan Goodman. emcee: The MCMC Hammer. *Publ. Astron. Soc. Pac.*, 125(925):306–312, mar 2013.
- [144] Bruce Allen, Warren G. Anderson, Patrick R. Brady, Duncan A. Brown, and Jolien D. E. Creighton. Findchirp: An algorithm for detection of gravitational waves from inspiraling compact binaries. *Phys. Rev. D*, 85:122006, Jun 2012.
- [145] Eric Thrane and Colm Talbot. An introduction to Bayesian inference in gravitational-wave astronomy: parameter estimation, model selection, and hierarchical models. *Publications of the Astronomical Society of Australia*, 36, 2019.
- [146] Cody Messick, Kent Blackburn, Patrick Brady, Patrick Brockill, Kipp Cannon, Romain Cariou, Sarah Caudill, Sydney J Chamberlin, Jolien DE Creighton, Ryan Everett, et al. Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data. *Phys. Rev. D*, 95(4):042001, 2017.
- [147] Chanseok Park Ayanendranath Basu, Hiroyuki Shioya. *Statistical Inference: The Minimum Distance Approach*. Chapman and Hall/CRC, 1st ed. edition, 2011.
- [148] Lalit Pathak, Sanket Munishwar, Amit Reza, and Anand S. Sengupta. Prompt sky localization of compact binary sources using meshfree approximation, 2023.
- [149] Michele Vallisneri. Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects. *Phys. Rev. D*, 77(4):042001, 2008.
- [150] C. Pankow, P. Brady, E. Ochsner, and R. O’Shaughnessy. Novel scheme for rapid parallel parameter estimation of gravitational waves from compact binary coalescences. *Phys. Rev. D*, 92:023002, Jul 2015.
- [151] Gian Mario Manca and Michele Vallisneri. Cover art: Issues in the metric-guided and metric-less placement of random and stochastic template banks. *Phys. Rev. D*, 81:024004, Jan 2010.
- [152] David Eichler, Mario Livio, Tsvi Piran, and David N. Schramm. Nucleosynthesis, Neutrino Bursts and Gamma-Rays from Coalescing Neutron Stars. *Nature*, 340:126–128, 1989.
- [153] J. Goodman. Are gamma-ray bursts optically thick? *apjl*, 308:L47, September 1986.
- [154] Brian D. Metzger. Kilonovae. *Living Reviews in Relativity*, 1(1):1, 2019.
- [155] Li-Xin Li and Bohdan Paczyński. Transient events from neutron star mergers. *The Astrophysical Journal*, 507(1):L59–L62, November 1998.

- [156] Ish Gupta, Ssohrab Borhanian, Arnab Dhani, Debatri Chattopadhyay, Rahul Kashyap, V. Ashley Villar, and B. S. Sathyaprakash. Neutron star-black hole mergers in next generation gravitational-wave observatories. *Phys. Rev. D*, 107:124007, Jun 2023.
- [157] Gregory M. Harry. Advanced LIGO: The next generation of gravitational wave detectors. *Class. Quant. Grav.*, 27:084006, 2010.
- [158] B. P. Abbott, R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, V. B. Adya, C. Affeldt, M. Agathos, et al. Prospects for observing and localizing gravitational-wave transients with advanced LIGO, advanced virgo and KAGRA. *Living Reviews in Relativity*, 23(1), sep 2020.
- [159] F Acernese, M Agathos, K Agatsuma, D Aisa, N Allemandou, A Allocata, J Amarni, P Astone, G Balestri, G Ballardin, et al. Advanced virgo: a second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, 32(2):024001, dec 2014.
- [160] J Aasi, B P Abbott, R Abbott, T Abbott, M R Abernathy, K Ackley, C Adams, T Adams, P Addesso, R X Adhikari, et al. Advanced LIGO. *Classical and Quantum Gravity*, 32(7):074001, mar 2015.
- [161] D. V. Martynov, E. D. Hall, B. P. Abbott, R. Abbott, T. D. Abbott, C. Adams, R. X. Adhikari, R. A. Anderson, S. B. Anderson, K. Arai, et al. Sensitivity of the advanced ligo detectors at the beginning of gravitational wave astronomy. *Phys. Rev. D*, 93:112004, Jun 2016.
- [162] Gavin P Lamb and Shiho Kobayashi. GRB 170817A as a jet counterpart to gravitational wave trigger GW 170817. *Mon. Notices of the Royal Astron. Soc.*, 478(1):733–740, 2018.
- [163] Amber K Lennon, Alexander H Nitz, and Duncan A Brown. Measuring the eccentricity of GW170817 and GW190425. *Monthly Notices of the Royal Astronomical Society*, 497(2):1966–1971, jul 2020.
- [164] Theocharis A. Apostolatos, Curt Cutler, Gerald J. Sussman, and Kip S. Thorne. Spin-induced orbital precession and its modulation of the gravitational waveforms from merging binaries. *Phys. Rev. D*, 49:6274–6297, Jun 1994.
- [165] Mark Hannam, Charlie Hoy, Jonathan E. Thompson, Stephen Fairhurst, Vivien Raymond, Marta Colleoni, Derek Davis, Héctor Estellés, Carl-Johan Haster, Adrian Helmling-Cornell, Sascha Husa, David Keitel, T. J. Massinger, Alexis Menéndez-Vázquez, Kentaro Mogushi, Serguei Ossokine, Ethan Payne, Geraint Pratten, Isobel Romero-Shaw, Jam Sadiq, Patricia Schmidt, Rodrigo Tenorio, Richard Udall, John Veitch, Daniel Williams, Anjali Balasaheb Yelikar, and Aaron Zimmerman. General-relativistic precession in a black-hole binary. *Nature*, 610(7933):652–655, October 2022.
- [166] Noah E. Wolfe, Salvatore Vitale, and Colm Talbot. Too small to fail: characterizing sub-solar mass black hole mergers with gravitational waves, 2023.

- [167] Antoni Ramos-Buades, Alessandra Buonanno, Héctor Estellés, Mohammed Khalil, Deyan P. Mihaylov, Serguei Ossokine, Lorenzo Pompili, and Mahlet Shiferaw. Seobnrv5phm: Next generation of accurate and efficient multipolar precessing-spin effective-one-body waveforms for binary black holes, 2023.
- [168] Tousif Islam, Javier Roulet, and Tejaswi Venumadhav. Factorized parameter estimation for real-time gravitational wave inference, 2022.
- [169] Chayan Chatterjee and Linqing Wen. Pre-merger sky localization of gravitational waves from binary neutron star mergers using deep learning, 2022.
- [170] Rory Smith, Scott E Field, Kent Blackburn, Carl-Johan Haster, Michael Pürrer, Vivien Raymond, and Patricia Schmidt. Fast and accurate inference on gravitational waves from precessing compact binaries. *Phys. Rev. D*, 94(4):044031, 2016.
- [171] Stephen Fairhurst, Charlie Hoy, Rhys Green, Cameron Mills, and Samantha A. Usman. Fast inference of binary merger properties using the information encoded in the gravitational-wave signal, 2023.
- [172] Carl Edward Rasmussen and Christopher KI Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.
- [173] Eunsub Lee, Soichiro Morisaki, and Hideyuki Tagoshi. Mass-spin reparametrization for a rapid parameter estimation of inspiral gravitational-wave signals. *Phys. Rev. D*, 105(12):124057, 2022.
- [174] Alvin J. K. Chua and Michele Vallisneri. Learning bayesian posteriors with neural networks for gravitational-wave inference. *Physical Review Letters*, 124(4), jan 2020.
- [175] Stephen R. Green, Christine Simpson, and Jonathan Gair. Gravitational-wave parameter estimation with autoregressive neural network flows. *Physical Review D*, 102(10), nov 2020.
- [176] Stephen R. Green and Jonathan Gair. Complete parameter inference for gw150914 using deep learning, 2020.
- [177] Kaze W. K. Wong, Maximiliano Isi, and Thomas D. P. Edwards. Fast gravitational wave parameter estimation without compromises, 2023.
- [178] Ronan Legin, Maximiliano Isi, Kaze Wong, Alexandre Adam, Laurence Perreault-Levasseur, and Yashar Hezaveh. Towards unbiased gravitational-wave parameter estimation using score-based likelihood characterization. In *2nd ICML Workshop on Machine Learning for Astrophysics*, Honolulu, Hawai'i, 2023.
- [179] Lalit Pathak, Amit Reza, and Anand S. Sengupta. Fast likelihood evaluation using meshfree approximations for reconstructing compact binary sources. *Phys. Rev. D*, 108:064055, Sep 2023.

- [180] Alex Nitz, Ian Harry, Duncan Brown, Christopher M. Biwer, Josh Willis, Tito Dal Canton, Collin Capano, Thomas Dent, Larne Pekowsky, Andrew R. Williamson, et al. gwastro/pycbc: v2.0.5 release of PyCBC, Jul 2022.
- [181] Vaibhav Tiwari, Charlie Hoy, Stephen Fairhurst, and Duncan MacLeod. Fast non-Markovian sampler for estimating gravitational-wave posteriors. *Phys. Rev. D*, 108:023001, Jul 2023.
- [182] J. Veitch, V. Raymond, B. Farr, W. Farr, P. Graff, S. Vitale, B. Aylott, K. Blackburn, N. Christensen, M. Coughlin, et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the lalinference software library. *Phys. Rev. D*, 91:042003, Feb 2015.
- [183] John T. Whelan. Notes on the geometry of gravitational wave detection. Technical Report LIGO-T1300666-v3, School of Gravitational Waves, Warsaw, July 2013. Notes for lectures at the Banach Centre School of Gravitational Waves, Warsaw, Poland, 2013 July 1-5.
- [184] C. Pankow, P. Brady, E. Ochsner, and R. O'Shaughnessy. Novel scheme for rapid parallel parameter estimation of gravitational waves from compact binary coalescences. *Phys. Rev. D*, 92:023002, Jul 2015.
- [185] Soumen Roy, Anand S. Sengupta, and Parameswaran Ajith. Effectual template banks for upcoming compact binary searches in advanced-ligo and virgo data. *Phys. Rev. D*, 99:024048, Jan 2019.
- [186] Debnandini Mukherjee, Sarah Caudill, Ryan Magee, Cody Messick, Stephen Privitera, Surabhi Sachdev, Kent Blackburn, Patrick Brady, Patrick Brockill, Kipp Cannon, et al. Template bank for spinning compact binary mergers in the second observation run of advanced ligo and the first observation run of advanced virgo. *Phys. Rev. D*, 103:084047, Apr 2021.
- [187] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Meth.*, 17:261–272, 2020.
- [188] LIGO Scientific Collaboration and Virgo Collaboration. Data release for event GW170817, 2017. Accessed on 2023-04-20.
- [189] Rich Abbott et al. Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo. *SoftwareX*, 13:100658, 2021.
- [190] Francesco Iacovelli, Michele Mancarella, Stefano Foffa, and Michele Maggiore. GWFEST: A fisher information matrix python code for third-generation gravitational-wave detectors. *The Astrophysical Journal Supplement Series*, 263(1):2, oct 2022.
- [191] Humberto Rocha. On the selection of the most adequate radial basis function. *Applied Mathematical Modelling*, 33(3):1573–1583, 2009.

- [192] Leo P. Singer and Larry R. Price. Rapid bayesian position reconstruction for gravitational-wave transients. *Phys. Rev. D*, 93:024013, Jan 2016.
- [193] Florian Gerber and Reinhard Furrer. optimParallel: An R Package Providing a Parallel Version of the L-BFGS-B Optimization Method. *The R Journal*, 11(1):352–358, 2019.
- [194] Florian Gerber. flora fauna/optimparallel-python: test zenodo, June 2020.
- [195] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, et al. Observation of gravitational waves from two neutron star–black hole coalescences. *The Astrophysical Journal Letters*, 915(1):L5, jun 2021.
- [196] Sabine Le Borne and Michael Wende. Domain decomposition methods in scattered data interpolation with conditionally positive definite radial basis functions. *Comput. Math. Appl.*, 77:1178–1196, 2019.
- [197] C. Pankow, P. Brady, E. Ochsner, and R. O’Shaughnessy. Novel scheme for rapid parallel parameter estimation of gravitational waves from compact binary coalescences. *Physical Review D*, 92(2), jul 2015.
- [198] Qi Zhang, Yangzhang Zhao, and Jeremy Levesley. Adaptive radial basis function interpolation using an error indicator. *Numerical Algorithms*, 76(2):441–471, 2017.
- [199] Walter Del Pozzo, Tjonne G. F. Li, Michalis Agathos, Chris Van Den Broeck, and Salvatore Vitale. Demonstrating the feasibility of probing the neutron star equation of state with second-generation gravitational wave detectors. *Phys. Rev. Lett.*, 111(7):071101, 2013.
- [200] Tim Dietrich, Anuradha Samajdar, Sebastian Khan, Nathan K. Johnson-McDaniel, Reetika Dudi, and Wolfgang Tichy. Improving the nrtidal model for binary neutron star systems. *Physical Review D*, 100(4), August 2019.
- [201] Harsh Narola, Justin Januart, Quirijn Meijer, K. Haris, and Chris Van Den Broeck. Relative binning for complete gravitational-wave parameter estimation with higher-order modes and precession, and applications to lensing and third-generation detectors, 2023.
- [202] L. Pathak. github repository of the relevant codes used in this paper, Dec 2023.
- [203] Tim Dietrich, Sebastian Khan, Reetika Dudi, Shasvath J. Kapadia, Prayush Kumar, Alessandro Nagar, Frank Ohme, Francesco Pannarale, Anuradha Samajdar, Sebastiano Bernuzzi, Gregorio Carullo, Walter Del Pozzo, Maria Haney, Charalampos Markakis, Michael Pürller, Gunnar Riemschneider, Yoshinta Eka Setyawati, Ka Wa Tsang, and Chris Van Den Broeck. Matter imprints in waveform models for neutron star binaries: Tidal and self-spin effects. *Phys. Rev. D*, 99:024029, Jan 2019.
- [204] Leslie Wade, Jolien D. E. Creighton, Evan Ochsner, Benjamin D. Lackey, Benjamin F. Farr, Tyson B. Littenberg, and Vivien Raymond. Systematic and statistical errors in a

bayesian approach to the estimation of the neutron-star equation of state using advanced gravitational wave detectors. *Phys. Rev. D*, 89:103012, May 2014.

- [205] R Abbott, TD Abbott, F Acernese, K Ackley, C Adams, N Adhikari, RX Adhikari, VB Adya, C Affeldt, D Agarwal, et al. Gwtc-3: compact binary coalescences observed by ligo and virgo during the second part of the third observing run. *arXiv preprint arXiv:2111.03606*, 2021.
- [206] R. Abbott, T. D. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. X. Adhikari, V. B. Adya, C. Affeldt, D. Agarwal, M. Agathos, et al. Observation of gravitational waves from two neutron star–black hole coalescences. *The Astrophysical Journal Letters*, 915(1):L5, jun 2021.
- [207] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, et al. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Phys. Rev. Lett.*, 119:161101, Oct 2017.
- [208] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, et al. Multi-messenger Observations of a Binary Neutron Star Merger. *The Astrophysical Journal*, 848(2):L12, oct 2017.
- [209] Adam G. Riess, Stefano Casertano, Wenlong Yuan, Lucas M. Macri, and Dan Scolnic. Large Magellanic Cloud Cepheid Standards Provide a 1% Foundation for the Determination of the Hubble Constant and Stronger Evidence for Physics beyond Λ CDM. *The Astrophysical Journal*, 876(1):85, may 2019.
- [210] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, et al. A gravitational-wave standard siren measurement of the Hubble constant. *Nature*, 551(7678):85–88, oct 2017.
- [211] Iair Arcavi. The first hours of the GW170817 kilonova and the importance of early optical and ultraviolet observations for constraining emission models. *The Astrophysical Journal*, 855(2):L23, mar 2018.
- [212] Doron Grossman, Oleg Korobkin, Stephan Rosswog, and Tsvi Piran. The long-term evolution of neutron star merger remnants – II. radioactively powered transients. *Monthly Notices of the Royal Astronomical Society*, 439(1):757–770, jan 2014.
- [213] L. F. Roberts, D. Kasen, W. H. Lee, and E. Ramirez-Ruiz. Electromagnetic transients powered by nuclear decay in the tidal tails of coalescing compact binaries. *The Astrophysical Journal Letters*, 736(1):L21, jul 2011.
- [214] B. D. Metzger, G. Martínez-Pinedo, S. Darbha, E. Quataert, A. Arcones, D. Kasen, R. Thomas, P. Nugent, I. V. Panov, and N. T. Zinner. Electromagnetic counterparts of compact object mergers powered by the radioactive decay of r-process nuclei. *Monthly Notices of the Royal Astronomical Society*, 406(4):2650–2662, jun 2010.

- [215] B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, et al. Gravitational Waves and Gamma-Rays from a Binary Neutron Star Merger: GW170817 and GRB 170817A. *The Astrophysical Journal*, 848(2):L13, oct 2017.
- [216] Jin-Jun Geng, Zi-Gao Dai, Yong-Feng Huang, Xue-Feng Wu, Long-Biao Li, Bing Li, and Yan-Zhi Meng. Brightening X-Ray/Optical/Radio Emission of GW170817/SGRB 170817A: Evidence for an Electron–Positron Wind from the Central Engine? *The Astrophysical Journal*, 856(2):L33, mar 2018.
- [217] G. Ghirlanda, O. S. Salafia, Z. Paragi, M. Giroletti, J. Yang, B. Marcote, J. Blanchard, I. Agudo, T. An, M. G. Bernardini, et al. Compact radio emission indicates a structured jet was produced by a binary neutron star merger. *Science*, 363(6430):968–971, mar 2019.
- [218] E. Troja, L. Piro, H. van Eerten, R. T. Wollaeger, M. Im, O. D. Fox, N. R. Butler, S. B. Cenko, T. Sakamoto, C. L. Fryer, et al. The X-ray counterpart to the gravitational-wave event GW170817. *Nature*, 551(7678):71–74, oct 2017.
- [219] Hsin-Yu Chen, Maya Fishbach, and Daniel E. Holz. A two per cent Hubble constant measurement from standard sirens within five years. *Nature*, 562(7728):545–547, oct 2018.
- [220] B. F. Schutz. Determining the Hubble constant from gravitational wave observations. *nature*, 323(6086):310–311, September 1986.
- [221] Andreas Finke, Stefano Foffa, Francesco Iacobelli, Michele Maggiore, and Michele Mancarella. Cosmology with LIGO/Virgo dark sirens: Hubble parameter and modified gravitational wave propagation. *Journal of Cosmology and Astroparticle Physics*, 2021(08):026, aug 2021.
- [222] M. Soares-Santos, A. Palmese, W. Hartley, J. Annis, J. Garcia-Bellido, O. Lahav, Z. Doctor, M. Fishbach, D. E. Holz, H. Lin, et al. First Measurement of the Hubble Constant from a Dark Standard Siren using the Dark Energy Survey Galaxies and the LIGO/Virgo Binary–Black-hole Merger GW170814. *The Astrophysical Journal*, 876(1):L7, apr 2019.
- [223] Leo P. Singer and Larry R. Price. Rapid bayesian position reconstruction for gravitational-wave transients. *Physical Review D*, 93(2), jan 2016.
- [224] Timothée Accadia, Fausto Acernese, F. Antonucci, P. Astone, G. Ballardin, Fabrizio Barone, M Barsuglia, A. Basti, Th Bauer, M Bebronne, Mark Beker, Arnaud Bellettoile, S Birindelli, Massimiliano Bitossi, M Bizouard, Marc Blom, François Bondu, Lucia Bonelli, R Bonnand, and J-P Zendri. Status of the Virgo project. *Classical and Quantum Gravity*, 28:114002, 05 2011.
- [225] Hsin-Yu Chen, Daniel E Holz, John Miller, Matthew Evans, Salvatore Vitale, and Jolien Creighton. Distance measures in gravitational-wave astrophysics and cosmology. *Classical and Quantum Gravity*, 38(5):055010, jan 2021.

- [226] The LIGO Scientific Collaboration, the Virgo Collaboration, J. Abadie, B. P. Abbott, R. Abbott, M Abernathy, T. Accadia, F. Acernese, C. Adams, R. Adhikari, et al. Sensitivity to Gravitational Waves from Compact Binary Coalescences Achieved during LIGO's Fifth and Virgo's First Science Run, 2010.
- [227] Lee Samuel Finn and David F. Chernoff. Observing binary inspiral in gravitational radiation: One interferometer. *Phys. Rev. D*, 47:2198–2219, Mar 1993.
- [228] Chris Pankow, Monica Rizzo, Kaushik Rao, Christopher P L Berry, and Vassiliki Kalogera. Localization of compact binary sources with second-generation gravitational-wave interferometer networks. *The Astrophysical Journal*, 902(1):71, oct 2020.
- [229] R. Weizmann Kiendrebeogo, Amanda M. Farah, Emily M. Foley, Abigail Gray, Nina Kunert, Anna Puecher, Andrew Toivonen, R. Oliver VandenBerg, Shreya Anand, Tomás Ahumada, et al. Updated observing scenarios and multi-messenger implications for the International Gravitational-wave Network's O4 and O5, 2023.
- [230] Surabhi Sachdev, Ryan Magee, Chad Hanna, Kipp Cannon, Leo Singer, Javed Rana SK, Debnandini Mukherjee, Sarah Caudill, Chiwai Chan, Jolien D. E. Creighton, et al. An early-warning system for electromagnetic follow-up of gravitational-wave events. *The Astrophysical Journal Letters*, 905(2):L25, dec 2020.
- [231] Ryan Magee, Deep Chatterjee, Leo P. Singer, Surabhi Sachdev, Manoj Kovalam, Geoffrey Mo, Stuart Anderson, Patrick Brady, Patrick Brockill, Kipp Cannon, Tito Dal Canton, et al. First demonstration of early warning gravitational-wave alerts. *The Astrophysical Journal Letters*, 910(2):L21, apr 2021.
- [232] Sylvia Biscoveanu, Salvatore Vitale, and Carl-Johan Haster. The reliability of the low-latency estimation of binary neutron star chirp mass. *Astrophys. J. Lett.*, 884(2):L32, 2019.
- [233] LVK. Noise curves used for simulations in the update of the observing scenarios paper. Technical Report T2000012-v2, LIGO-Virgo-KAGRA Scientific Collaboration, 2022.
- [234] T. Akutsu, M. Ando, K. Arai, Y. Arai, S. Araki, A. Araya, N. Aritomi, Y. Aso, S. W. Bae, Y. B. Bae, et al. Overview of KAGRA: Detector design and construction history, 2020.
- [235] Bernard F Schutz. Networks of gravitational wave detectors and three figures of merit. *Classical and Quantum Gravity*, 28(12):125023, may 2011.
- [236] M Saleem, Javed Rana, V Gayathri, Aditya Vijaykumar, Srashti Goyal, Surabhi Sachdev, Jishnu Suresh, S Sudhagar, Arunava Mukherjee, Gurudatt Gaur, et al. The science case for LIGO-India. *Classical and Quantum Gravity*, 39(2):025004, dec 2021.
- [237] Stephen Fairhurst. Source localization with an advanced gravitational wave detector network. *Classical and Quantum Gravity*, 28(10):105021, apr 2011.

- [238] Linqing Wen and Yanbei Chen. Geometrical expression for the angular resolution of a network of gravitational-wave detectors. *Physical Review D*, 81(8), apr 2010.
- [239] Chris Pankow, Eve A. Chase, Scott Coughlin, Michael Zevin, and Vassiliki Kalogera. Improvements in Gravitational-wave Sky Localization with Expanded Networks of Interferometers. *The Astrophysical Journal*, 854(2):L25, feb 2018.
- [240] Hsin-Yu Chen and Daniel E. Holz. Facilitating Follow-up of LIGO–Virgo Events Using Rapid Sky Localization. *The Astrophysical Journal*, 840(2):88, may 2017.
- [241] Leo P. Singer, Larry R. Price, Ben Farr, Alex L. Urban, Chris Pankow, Salvatore Vitale, John Veitch, Will M. Farr, Chad Hanna, Kipp Cannon, Tom Downes, Philip Graff, Carl-Johan Haster, Ilya Mandel, Trevor Sidery, and Alberto Vecchio. THE FIRST TWO YEARS OF ELECTROMAGNETIC FOLLOW-UP WITH ADVANCED LIGO AND VIRGO. *The Astrophysical Journal*, 795(2):105, oct 2014.
- [242] Carl L. Rodriguez, Benjamin Farr, Vivien Raymond, Will M. Farr, Tyson B. Littenberg, Diego Fazi, and Vicky Kalogera. BASIC PARAMETER ESTIMATION OF BINARY NEUTRON STAR SYSTEMS BY THE ADVANCED LIGO/VIRGO NETWORK. *The Astrophysical Journal*, 784(2):119, mar 2014.
- [243] Lalit Pathak, Amit Reza, and Anand S. Sengupta. Fast likelihood evaluation using meshfree approximations for reconstructing compact binary sources. *Phys. Rev. D*, 108:064055, Sep 2023.
- [244] Vishal Baibhav, Emanuele Berti, Davide Gerosa, Michela Mapelli, Nicola Giacobbo, Yann Bouffanais, and Ugo N. Di Carlo. Gravitational-wave detection rates for compact binaries formed in isolation: Ligo/virgo o3 and beyond. *Physical Review D*, 100(6), September 2019.
- [245] I M Romero-Shaw, C Talbot, S Biscoveanu, V D’Emilio, G Ashton, C P L Berry, S Coughlin, S Galauadage, C Hoy, M Hübner, et al. Bayesian inference for compact binary coalescences with bilby: Validation and application to the first ligo–virgo gravitational-wave transient catalogue. *Monthly Notices of the Royal Astronomical Society*, 499(3):3295–3319, sep 2020.
- [246] Gregory Ashton, Moritz Hübner, Paul D. Lasky, Colm Talbot, Kendall Ackley, Sylvia Biscoveanu, Qi Chu, Atul Divakarla, Paul J. Easter, Boris Goncharov, Francisco Hernandez Vivanco, et al. Bilby: A user-friendly bayesian inference library for gravitational-wave astronomy. *The Astrophysical Journal Supplement Series*, 241(2):27, apr 2019.
- [247] Geraint Pratten, Cecilio García-Quirós, Marta Colleoni, Antoni Ramos-Buades, Héctor Estellés, Maite Mateu-Lucena, Rafel Jaume, Maria Haney, David Keitel, Jonathan E. Thompson, and Sascha Husa. Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes. *Physical Review D*, 103(10), May 2021.

- [248] Kipp Cannon, Sarah Caudill, Chiwai Chan, Bryce Cousins, Jolien DE Creighton, Becca Ewing, Heather Fong, Patrick Godwin, Chad Hanna, Shaun Hooper, et al. Gstlal: A software framework for gravitational wave discovery. *SoftwareX*, 14:100680, 2021.
- [249] Harsh Kumar, Varun Bhalerao, G. C. Anupama, Sudhanshu Barway, Judhajeet Basu, Kunal Deshmukh, Kishalay De, Anirban Dutta, Christoffer Fremling, Hrishikesh Iyer, et al. India’s First Robotic Eye for Time-domain Astrophysics: The GROWTH-India Telescope. *The Astronomical Journal*, 164(3):90, aug 2022.
- [250] D. M. Macleod, J. S. Areeda, S. B. Coughlin, T. J. Massinger, and A. L. Urban. GWpy: A Python package for gravitational-wave astrophysics. *SoftwareX*, 13:100657, 2021.
- [251] Martin Beroiz and Jonah Kanner. Accessing Data with The Gravitational Wave Open Science Center. In *American Astronomical Society Meeting Abstracts*, volume 54 of *American Astronomical Society Meeting Abstracts*, page 348.09, June 2022.
- [252] I. Ebrová, M. Bílek, M. K. Yıldız, and J. Eliášek. NGC 4993, the shell galaxy host of GW170817: constraints on the recent galactic merger. *Astron. Astrophys.*, 634:A73, 2020.
- [253] Jens Hjorth, Andrew J. Levan, Nial R. Tanvir, Joe D. Lyman, Radoslaw Wojtak, Sophie L. Schroder, Ilya Mandel, Christa Gall, and Sofie H. Bruun. The distance to NGC 4993: The host galaxy of the gravitational-wave event GW170817. *The Astrophysical Journal*, 848(2):L31, oct 2017.
- [254] LIGO Scientific Collaboration and Virgo Collaboration. GWTC-2.1: Deep Extended Catalog of Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run - Parameter Estimation Data Release, may 2022.
- [255] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — Parameter estimation data release, November 2021.
- [256] S. A. Usman, J. C. Mills, and S. Fairhurst. Constraining the inclinations of binary mergers from gravitational-wave observations. *The Astrophysical Journal*, 877(2):82, May 2019.
- [257] Daniel Finstad, Soumi De, Duncan A. Brown, Edo Berger, and Christopher M. Biwer. Measuring the viewing angle of GW170817 with electromagnetic and gravitational waves. *The Astrophysical Journal*, 860(1):L2, jun 2018.
- [258] Christopher P. L. Berry, Ilya Mandel, Hannah Middleton, Leo P. Singer, Alex L. Urban, Alberto Vecchio, Salvatore Vitale, Kipp Cannon, Ben Farr, Will M. Farr, et al. PARAMETER ESTIMATION FOR BINARY NEUTRON-STAR COALESCENCES WITH REALISTIC NOISE DURING THE ADVANCED LIGO ERA. *The Astrophysical Journal*, 804(2):114, may 2015.
- [259] Anuradha Samajdar, Justin Janquart, Chris Van Den Broeck, and Tim Dietrich. Biases in parameter estimation from overlapping gravitational-wave signals in the third-generation detector era. *Physical Review D*, 104(4), aug 2021.

- [260] Justin Janquart, Tomasz Baka, Anuradha Samajdar, Tim Dietrich, and Chris Van Den Broeck. Analyses of overlapping gravitational wave signals using hierarchical subtraction and joint parameter estimation. *Mon. Not. Roy. Astron. Soc.*, 523(2):1699–1710, 2023.
- [261] James Alvey, Uddipta Bhardwaj, Samaya Nissanke, and Christoph Weniger. What to do when things get crowded? scalable joint analysis of overlapping gravitational wave signals, 2023.