Towards high-precision gravitational wave astronomy: robust and efficient data analysis for ground-based detectors

Qian Hu

Submitted in fulfilment of the requirements for the
Degree of Doctor of Philosophy

School of Physics and Astronomy
College of Science and Engineering
University of Glasgow

October 2024

Abstract

Since the first direct detection of Gravitational Waves (GWs) by Laser Interferometer Gravitational-Wave Observatory (LIGO), GW astronomy has advanced rapidly across theoretical, observational, and instrumental sciences. While the detection rate from the LIGO-Virgo-KAGRA collaboration (LVK) continues to rise, several Third-Generation (3G) ground-based GW detectors are being proposed for the 2030s, aiming to detect millions of GW events per year with significantly improved signal-to-noise ratios. The increased precision in GW astronomy will generate vast amounts of data, posing challenges in data analysis concerning robustness and efficiency. Ensuring robustness in data analysis is crucial for deriving accurate scientific conclusions, while efficiency is essential for performing analyses within manageable timeframes and hardware constraints—especially for time-sensitive tasks in transient astronomy.

This thesis aims to investigate the challenges of robustness and efficiency and explore possible solutions. In Chapters 1 and 2, we give an overview of the basic concepts in GW astrophysics and data analysis, and bring up the robustness and efficiency challenges. For robustness, we show how data analysis can lead to incorrect scientific conclusions using the example of testing general relativity with inaccurate waveforms and overlapped signals in Chapter 3. We investigate the error accumulation effects on the catalog level and we identify the waveform inaccuracy as the primary contributor to systematic errors. Following this, in Chapter 4, we propose a waveform accuracy assessment approach that can be readily applied to parameter estimation results without numerical relativity simulations. With this method, we examine parameter estimation results from the latest LVK public event catalogs GWTC-2.1 and GWTC-3, and make predictions of waveform accuracy requirements for the future detectors. For efficiency, we focus on the long Binary Neutron Star (BNS) signals expected in the 3G detectors. In Chapter 5, we demonstrate pre-merger source localization for long BNS signals with multi-band matched filtering and a semi-analytical localization algorithm. Using our method, we show that it is possible to provide accurate sky localizations more than 30 minutes before the merger. We also provide a forecast on the detection rate of well-localized early-warning BNS events. Further, in Chapter 6, we develop machine learning models equipped with a suite of data preprocessing methods for the full parameter estimation of hours-long BNS signals, which is

prohibitively slow using traditional methods. The models' precision is validated against analytical forecasts and the accuracy is confirmed by self-consistency tests. The thesis concludes with a summary of the findings and an outlook on high-precision GW data analysis in Chapter 7.

Contents

List of Tables

List of Figures

Acknowledgements

Time flies. As I reach this milestone, I might have to first thank my friend Tianyi Zhang, who shared numerous intriguing physics books that lead me to this path in middle school, who also shared the excitements of LIGO's first detection on a bus to a bookstore on a Friday. Despite carrying the youthful enthusiasm for science from my teenage years, the reality of scientific research has often drained that passion. I must be grateful to the many people who have supported, encouraged, and brought joy to me throughout my PhD journey.

My heartfelt thanks go to all members of the Institute for Gravitational Research (IGR), whose collaboration and support have been invaluable throughout this process. The welcoming and relaxing environment made my time at IGR enjoyable and rewarding.

My research has greatly benefited from insightful discussions and generous help from my colleagues. Dr. Daniel Williams assisted me with accessing the LIGO computing clusters and GWTC data, greatly simplifying my work in waveform accuracy assessment (Chapter 4). Dr. Chayan Chatterjee and Prof. Linqing Wen provided helpful discussions when I was struggling with the bugs in `SealGW`, and Daniel Tang taught me how to use `git` properly during the development of the `SealGW` Python package (Chapter 5). I am thankful to Qi Sun and Xueting Zhang for many useful discussions on machine learning while I was new to the field. Maximilian Dax and Nihar Gupte and many other `DINGO` group members generously shared me with their expertise in neural density estimation. These all lead to my results in Chapter 6. Dr. Chris Messenger also provided insightful comments for Chapter 6. Dr. Yifan Wang offered generous help during my job applications, and shared many interesting ideas throughout my PhD. Xiaolin Liu helped me with his eccentric waveform model. While it is impossible to thank everyone individually, I am deeply grateful to all who have supported me and contributed to my research.

I am indebted to my supervisors. Prof. Ik Siong Heng and his family have always been welcoming, making me feel at home in Glasgow. I am grateful to their hospitality and delicious dinners. Dr. John Veitch is always supportive, patient, and encouraging to me. I appreciate the time he spent sitting with me, helping me with coding, debugging, and analyzing results. If I ever have the opportunity to be a supervisor, I hope I can be as kind and supportive as they have been.

Finally, I am eternally grateful to my parents for their unwavering support and love, for the way they educated me, and for always believing in me.

My study was financially supported by the China Scholarship Council and Lord Kelvin / Charles Lindie Mitchell travel fund. I made extensive use of the computational resources provided by the LIGO Lab at Caltech, IGR and Cardiff University.

Declaration

This dissertation is the result of my own work under the supervision of Dr. John Veitch in the Institute for Gravitational Research at the University of Glasgow. It has not been submitted for any other degree at the University of Glasgow or any other institution. However, many of the results presented have been published as journal articles. These are:

- Qian Hu and John Veitch. Accumulating Errors in Tests of General Relativity with Gravitational Waves: Overlapping Signals and Inaccurate Waveforms. *Astrophys. J., 945(2) 103, 2023* [1]. This is presented in Chapter 3.
- Qian Hu and John Veitch. Assessing the model waveform accuracy of gravitational waves. *Phys. Rev. D, 106(4) 044042, 2022* [2]. This is presented in Chapter 4.
- Qian Hu and John Veitch. Rapid Premerger Localization of Binary Neutron Stars in Third-generation Gravitational-wave Detectors. *Astrophys. J. Lett., 958(2) L43, 2023* [3]. This is presented in Chapter 5. Note that the `SealGW` algorithm used in this chapter is my work during my undergraduate study at University of Science and Technology of China and was published in *Physical Review D* [4]. I briefly introduce `SealGW` in Sec. 5.2 for completeness, but the algorithm itself is not the primary result I intend to show in Chapter 5. I also note that the Python package for `SealGW` was developed during my PhD study at the University of Glasgow.

Results presented in Chapter 6 are not published in any journal articles by the time of writing this dissertation.

This dissertation contains ~ 44000 words, counted by VSCode LaTeX Utilities.

Notations and Conventions

Indices

I will use Greek letters $(\mu, \nu, \dots)$ for tensor indices in four-dimension spacetime and Latin letters $(i, j, k, \dots)$ for the three-dimension spatial indices and matrix indices. 0 is used for the time-like index for four-dimension tensors. Repeated indices represent summation unless otherwise stated.

Metric

The generic metric is denoted as $g_{\mu\nu}$; Minkowski metric is denoted as $\eta_{\mu\nu}$ and has sign $(-,+,+,+)$.

Unit and constants

By default, I set the speed of light $c = 1$ and write the Newton constant G explicitly. I may use geometrized units, $c = G = 1$, in contexts related to gravitational wave waveforms and the use will be stated in advance. The solar mass is denoted by $M_\odot$.

Riemann and Ricci tensor, Einstein equations

The conventions for Riemann and Ricci tensor and Einstein equations are aligned with Misner, Thorne and Wheeler (1973) [5]. The Christoffel symbol is

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

and the Riemann tensor is defined as

$$R^{\mu}{}_{\nu\rho\sigma} = \partial_\rho \Gamma^{\mu}_{\nu\sigma} - \partial_\sigma \Gamma^{\mu}_{\nu\rho} + \Gamma^{\mu}_{\alpha\rho} \Gamma^{\alpha}_{\nu\sigma} - \Gamma^{\mu}_{\alpha\sigma} \Gamma^{\alpha}_{\nu\rho}. \tag{2}$$

The Ricci tensor is $R_{\mu\nu} = R^{\alpha}_{\mu\alpha\nu}$ and Ricci scalar is $R = g^{\mu\nu} R_{\mu\nu}$. The Einstein equations (or, Einstein field equations) read

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{3}$$

where $T_{\mu\nu}$ is the energy-momentum tensor.

Fourier transform in data analysis

The Fourier transform for the time-domain data $d(t)$ is

$$\tilde{d}(f) = \int_{-\infty}^{+\infty} d(t) e^{-2\pi i f t} dt, \tag{4}$$

and the inverse Fourier transform is

$$d(t) = \int_{-\infty}^{+\infty} \tilde{d}(f) e^{2\pi i f t} df. \tag{5}$$

I use *tilde* to denote frequency domain data when there might be confusion. I may also omit the *tilde* when there is no necessity to distinguish frequency domain signals and time domain signals.

Acronyms

2G Second-Generation. 40

3G Third-Generation. i, 40, 52, 73, 74, 75, 76, 77, 78, 94, 116, 117, 125, 128, 131, 136, 139, 140, 141, 148, 152, 164, 167, 168, 171, 172

AGN Active galactic nuclei. 22

AI Artificial Intelligence. 68

aLIGO Advanced Laser Interferometer Gravitational-Wave Observatory. 26, 35, 39

ASD Amplitude spectral Density. 35

BBH Binary Black Hole. 11, 21, 22, 39, 47, 48, 50, 72, 82, 83, 84, 85, 86, 88, 89, 91, 92, 95, 101, 102, 103, 104, 109, 110, 111, 114, 128, 147, 148, 149

BH Black Hole. 22

BNS Binary Neutron Star. i, 22, 23, 26, 39, 44, 47, 75, 76, 85, 86, 95, 101, 110, 112, 113, 115, 116, 117, 128, 129, 131, 133, 134, 135, 136, 139, 140, 148, 149, 150, 151, 152, 153, 154, 155, 159, 161, 162, 163, 164, 167, 168, 170, 171

CBC Compact Binary Coalescence. 1, 23, 24, 25, 27, 39, 40, 42, 46, 47, 49, 52, 54, 62, 63, 72, 73, 75, 77, 78, 85, 109, 128, 141, 148, 150, 169

CE Cosmic Explorer. 40, 117, 131, 156, 158, 168, 171

CMB Cosmic Microwave Background. 25

CNN Convolutional Neural Network. 70, 71, 147

CPTA Chinese Pulsar Timing Array. 42

DECIGO DECI-hertz Gravitational-wave Observatory. 41

DWD Double White Dwarf. 22

EM Electromagnetic. 23, 24, 25, 39, 72, 76, 117, 128, 136

EMRI Extreme-Mass-Ratio Inspiral. 22, 23, 40

EOB Effective-One-Body. 49, 50, 51, 52, 73

EPTA European Pulsar Timing Array. 42

ESA European Space Agency. 40, 41

ET Einstein Telescope. 40, 117, 131, 133, 156, 158, 168, 171

FAR False Alarm Rate. 62

GAN Generative Adversarial Networks. 72, 73

GAST Greenwich Apparent Sidereal Time. 155

GEO-600 GEO-600 Gravitational Wave Detector. 39

GMST Greenwich Mean Sidereal Time. 31, 155, 158

GR General Relativity. 1, 2, 10, 15, 20, 21, 26, 46, 47, 54, 76, 77, 78, 80, 81, 83, 85, 87, 88, 89, 90, 91, 92, 94, 169, 170

GRB Gamma-Ray Burst. 23, 25, 39

GW Gravitational Wave. i, ii, 1, 2, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 33, 34, 35, 39, 40, 41, 42, 43, 46, 47, 48, 49, 52, 53, 54, 55, 56, 62, 66, 67, 68, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 83, 85, 86, 91, 94, 95, 96, 100, 101, 102, 105, 106, 108, 109, 114, 115, 116, 117, 118, 121, 125, 128, 129, 130, 131, 133, 136, 141, 145, 146, 147, 150, 151, 152, 153, 156, 164, 169, 170, 171, 172

GWTC Gravitational-Wave Transient Catalog. 101

GWTC-2 Gravitational-Wave Transient Catalog 2. 102, 103, 104

GWTC-1 Gravitational-Wave Transient Catalog 1. 101, 103

GWTC-3 Gravitational-Wave Transient Catalog 3. i, 91, 95, 101, 102, 104, 106, 108, 110, 112, 114, 170

GWTC-2.1 Gravitational-Wave Transient Catalog 2.1. i, 95, 101, 102, 103, 104, 106, 107, 108, 110, 112, 114, 115, 170

H1 LIGO Detector at Hanford. 39, 131

IMRPhenom Inspiral-Merger-Ringdown phenomenological. 49, 51, 52

IPTA International Pulsar Timing Array. 42

ISCO Innermost Stable Circular Orbit. 20, 22, 46

JSD Jensen-Shannon Divergence. 65, 66

KAGRA Kamioka Gravitational Wave Detector. 39

KLD Kullback-Leibler Divergence. 65, 71, 145, 146

KS Test Kolmogorov-Smirnov Test. 66

L1 LIGO Detector at Livingston. 39, 131

LIGO Laser Interferometer Gravitational-Wave Observatory. i, 11, 15, 26, 27, 28, 35, 86, 111

LISA Laser Interferometer Space Antenna. 40

LVK LIGO-Virgo-KAGRA collaboration. i, 39, 72, 75, 76, 101, 117, 118

MCMC Markov Chain Monte Carlo. 63, 64, 65

ML Machine Learning. 68, 69, 71, 72, 73, 76

MLP Multi-Layer Perceptron. 70, 71, 72, 144, 145, 147, 148, 158

NANOGrav North American Nanohertz Observatory for Gravitational Waves. 42

NN Neural Network. 69, 70, 71, 72, 73

NR Numerical Relativity. 47, 48, 49, 50, 51, 52, 54, 95, 96, 97, 106, 110, 114, 115, 169

NRSur Numerical Relativity Surrogate. 50, 52

NS Neutron Star. 22

NSBH Neutron Star Black Hole Binaries. 22, 95, 101, 102, 110, 112, 113, 114, 128, 170

NSF Neural Spline Flow. 144, 145

O1 The First Observing Run of LVK. 39

O2 The Second Observing Run of LVK. 28, 39, 121, 122, 125, 127

O3 The Third Observing Run of LVK. 39, 127

O4 The Fourth Observing Run of LVK. 28, 39, 62, 72, 115, 128, 170

PE Parameter Estimation. 62, 74, 78, 79, 81, 92, 154, 170

PN Post-Newtonian. 46, 47, 49, 52, 81, 152

PPN Parametrized Post-Newtonian. 77, 81, 83, 93

PPTA Parkes Pulsar Timing Array. 42

PSD Power spectral Density. 35, 36, 56, 86

PTA Pulsar Timing Array. 27, 42

PyPI Python Package Index. 127

QFT Quantum Field Theory. 15

QNM Quasi-Normal Mode. 22, 48, 49

RealNVP Real Non-Volume Preserving. 144

ReLU Rectified Linear Unit. 70

ResNet Residual Network. 70, 144

ROM Reduced-Order Modeling. 49, 50, 51, 52

SealGW Semi-Analytical localization for Gravitational Waves. 116, 117, 118, 127, 128, 131, 132, 133, 136, 138, 159, 161, 163, 164, 168, 170, 171

SGWB Stochastic Gravitational-Wave Background. 23

SI International System of Units. 11, 18

SMBBH Super-Massive Binary Black Hole. 22, 23, 40, 42

SNR Signal-to-Noise Ratio. 28, 39, 57, 58, 59, 60, 61, 68, 73, 75, 77, 78, 79, 83, 84, 85, 88, 89, 91, 92, 96, 97, 100, 101, 102, 103, 105, 106, 108, 109, 110, 111, 112, 113, 114, 115, 120, 121, 122, 123, 124, 125, 126, 128, 130, 131, 133, 135, 140, 141, 154, 155, 157, 161, 162, 168, 169, 171

SPA Stationary Phase Approximation. 19, 46

SVD Singular Value Decomposition. 49, 61, 146, 147, 148, 149, 151, 152, 153, 154, 156, 157, 158, 164, 171

TT gauge Transverse-Traceless gauge. 6, 10, 11, 12, 14, 28

V1 VIRGO Detector. 39

ViT Vision Transformer. 148, 158, 164, 167

WD White Dwarf. 22

Chapter 1

Gravitational waves: physics, sources & detectors

Gravitational Waves (GWs) are ripples of spacetime predicted by Einstein's General Relativity (GR) in 1916, revealing that perturbations of spacetime caused by massive accelerating objects can propagate as waves. Theoretical investigations on GWs not only enrich the picture of the physical universe but also provide valuable insights and tools into observational astrophysics. This chapter gives a brief introduction to the physics of GWs. I will begin with the Einstein field equations and linear expansion to derive the basic properties of GWs in Sec. 1.1. Next, I will discuss the astrophysical sources of GWs in Sec. 1.2 and the methods for detecting them in Sec. 1.3. In Sec. 1.4, I will introduce waveform modeling for Compact Binary Coalescence (CBC) signals, which will be the focus of this thesis.

1.1 Gravitational waves under general relativity

The concept of Gravitational Waves (GWs) (the gravitational equivalent of electromagnetic waves) can be traced back to 1893, first proposed by Oliver Heaviside. However, it was not theorized until 1916 when Albert Einstein predicted the existence of gravitational radiation in spacetime, i.e. GWs, with his General Relativity (GR), in which spacetime is described by Riemann manifold and gravity is essentially the curvature of the manifold. Although there was initial skepticism about the existence of GWs, even from Einstein himself, due to gauge symmetry and singularity issues, it was finally proven that GW is a well-defined concept mathematically and physically in the 1960s. The theory of GWs rapidly advanced thereafter, including GWs in many non-GR theories. Given the breadth of the subject, this section will introduce the basic theory of GWs, focusing on the linearized theory within the framework of GR, and will discuss the properties of GWs. The equations and conventions for general relativity closely follow the textbook by Maggiore [6], except that we set $c = 1$.

1.1.1 The Einstein field equations and linearized theory

In general relativity, the geometry of spacetime and the distribution of matter are related via the Einstein field equations

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{1.1}$$

where $R_{\mu\nu}$ is the Ricci tensor, R is the Ricci scalar, $g_{\mu\nu}$ is the metric of the spacetime and $T_{\mu\nu}$ is the energy-momentum tensor (see detailed definitions in Notations and Conventions). The left-hand side, often referred to as the Einstein tensor $G_{\mu\nu}$, describes how spacetime is curved; the right-hand side contains the mass (energy), momentum and stress within that spacetime. The Einstein's equations reveal the essence of gravity to a higher level than Newton's theory does: gravity originates from the curvature of spacetime and tells matter how to move, while matter, in the meantime, dictates how spacetime curves.

The Einstein field equations are tensor equations relating a set of $(0, 2)$ tensors in the four-dimensional spacetime. When fully written out, the equations are 16 coupled and nonlinear differential equations, but not all equations are independent. By construction, $R_{\mu\nu}$, $g_{\mu\nu}$ and $T_{\mu\nu}$ are all symmetric tensors (e.g. $R_{\mu\nu} = R_{\nu\mu}$), which naturally reduces the number of independent components to 10. Some of the remaining degrees of freedom could be fixed by gauge symmetry. General relativity is invariant under all possible coordinate transformations $x^\mu \to x'^\mu$, where x'^μ can be any differentiable function of x^μ that has a differentiable inverse. This allows us to choose convenient coordinate systems to simplify the Einstein's equations in different scenarios, as we will do later in the linear expansion

form of the Einstein's equations.

Consider a perturbation in a flat spacetime, the metric can be expanded as

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

where $\eta_{\mu\nu}$ is the Minkowski metric of flat spacetime and $h_{\mu\nu}$ is the perturbation to the spacetime. Since any derivatives of $\eta_{\mu\nu}$ are zero, the Christoffel symbols and Riemann tensor can be easily calculated to the linear order (i.e. ignoring $\mathcal{O}(h^2)$ terms)

$$\begin{aligned} \Gamma^{\rho}_{\mu\nu} &= \frac{1}{2} g^{\rho\lambda} \left(\partial_\mu g_{\nu\lambda} + \partial_\nu g_{\lambda\mu} - \partial_\lambda g_{\mu\nu} \right) \\ &= \frac{1}{2} \eta^{\rho\lambda} \left(\partial_\mu h_{\nu\lambda} + \partial_\nu h_{\lambda\mu} - \partial_\lambda h_{\mu\nu} \right), \end{aligned} \tag{1.3}$$

$$\begin{aligned} R_{\mu\nu\rho\sigma} &= \eta_{\mu\lambda} \partial_\rho \Gamma^{\lambda}_{\nu\sigma} - \eta_{\mu\lambda} \partial_\sigma \Gamma^{\lambda}_{\nu\rho} \\ &= \frac{1}{2} \left(\partial_\rho \partial_\nu h_{\mu\sigma} + \partial_\sigma \partial_\mu h_{\nu\rho} - \partial_\sigma \partial_\nu h_{\mu\rho} - \partial_\rho \partial_\mu h_{\nu\sigma} \right). \end{aligned} \tag{1.4}$$

Contracting over μ and ρ we can get the Ricci tensor $R_{\mu\nu}$, further contracting μ and ν we can get the Ricci scalar R:

$$R_{\mu\nu} = \frac{1}{2} \left(\partial_\sigma \partial_\nu h^{\sigma}_{\mu} + \partial_\sigma \partial_\mu h^{\sigma}_{\nu} - \partial_\mu \partial_\nu h - \Box h_{\mu\nu} \right), \tag{1.5}$$

$$R = \partial_\mu \partial_\nu h^{\mu\nu} - \Box h. \tag{1.6}$$

where $\Box = -\partial_0^2 + \partial_1^2 + \partial_2^2 + \partial_3^2 = -\partial_0^2 + \nabla^2$ is the d'Alembert operator. Putting them together, we can write the Einstein tensor, i.e. the right-hand side of the Einstein field equations to the linear order under perturbation Eq. 1.2

$$\begin{aligned} G_{\mu\nu} &= R_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} R \\ &= \frac{1}{2} \left(\partial_\sigma \partial_\nu h^{\sigma}{}_{\mu} + \partial_\sigma \partial_\mu h^{\sigma}{}_{\nu} - \partial_\mu \partial_\nu h - \Box h_{\mu\nu} - \eta_{\mu\nu} \partial_\mu \partial_\nu h^{\mu\nu} + \eta_{\mu\nu} \Box h \right). \end{aligned} \tag{1.7}$$

The above expression can be further simplified by defining the trace-reversed part of $h_{\mu\nu}$, similar to the relation between the Ricci tensor and the Einstein tensor:

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2} \eta_{\mu\nu} h, \tag{1.8}$$

where $h = \eta^{\mu\nu} h_{\mu\nu}$ is the trace of $h_{\mu\nu}$. The trace of $\bar{h}_{\mu\nu}$ equals $-h$, and the same transformation to $\bar{h}_{\mu\nu}$ can bring $h_{\mu\nu}$ back. The Einstein tensor can be written in terms of $\bar{h}_{\mu\nu}$ as follows:

$$G_{\mu\nu} = -\frac{1}{2} \left(\Box \bar{h}_{\mu\nu} + \eta_{\mu\nu} \partial^\rho \partial^\sigma \bar{h}_{\rho\sigma} - \partial^\rho \partial_\nu \bar{h}_{\mu\rho} - \partial^\rho \partial_\mu \bar{h}_{\nu\rho} \right). \tag{1.9}$$

As mentioned earlier, general relativity is invariant under all possible smooth coordi-

nate transformations. We want to find a set of coordinates that simplifies Eq. 1.9. Intuitively, if $\bar{h}_{\mu\nu}$ follows the Lorenz gauge condition (in analogy to Lorenz gauge condition in electrodynamics $\partial_\mu A^\mu = 0$)

$$\partial^\mu \bar{h}_{\mu\nu} = 0, \tag{1.10}$$ [1]

then the last 3 terms in Eq. 1.9 will vanish. We can prove that this is possible. Consider the following coordinate transformation

$$x^\mu \to x'^\mu = x^\mu + \xi^\mu(x). \tag{1.11}$$

The metric will transform as

$$g_{\mu\nu}(x) \to g'_{\mu\nu}(x') = \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\sigma}{\partial x'^\nu} g_{\rho\sigma}(x). \tag{1.12}$$

Assuming derivatives of ξ^μ are of the same order of $h_{\mu\nu}$ and substituting Eq. 1.2 and Eq. 1.8, $\bar{h}_{\mu\nu}$ is transformed to

$$\bar{h}_{\mu\nu} \to \bar{h}'_{\mu\nu} = \bar{h}_{\mu\nu} - \left(\partial_\mu \xi_\nu + \partial_\nu \xi_\mu - \eta_{\mu\nu}\partial_\rho \xi^\rho\right), \tag{1.13}$$

and therefore

$$\partial^\nu \bar{h}_{\mu\nu} \to \left(\partial^\nu \bar{h}_{\mu\nu}\right)' = \partial^\nu \bar{h}_{\mu\nu} - \Box \xi_\mu. \tag{1.14}$$

This means that, if Lorenz gauge condition is not satisfied under the current coordinate, i.e., $\partial^\nu \bar{h}_{\mu\nu} = f_\mu(x) \neq 0$, we can always find a new coordinate by transformation Eq. 1.11 such that

$$\Box \xi_\mu = f_\mu(x), \tag{1.15}$$

which admits solutions since the d'Alembert operator is invertible, and therefore Lorenz gauge condition is satisfied in the new coordinate. In other words, we can always adopt the Lorenz gauge condition. The solution to Eq. 1.15 is given by the Green's function of the d'Alembert operator

$$\xi_\mu(x) = \int d^4x G(x-y) f_\mu(y), \tag{1.16}$$

where

$$\Box_x G(x-y) = \delta^4(x-y). \tag{1.17}$$

Referring back to the field equations, since the last three terms in Eq. 1.9 vanish under the Lorenz gauge condition, the field equations with linear perturbation are simplified to

[1] This is also called the Hilbert gauge, the Harmonic gauge or the De Donder gauge as they are essentially equivalent in the flat background. The latter two are more generic, defined in curved background as $\partial_\mu\left(g^{\mu\nu}\sqrt{-g}\right) = 0$. The name "Lorenz" is sometimes misspelled as "Lorentz" in some literature - the gauge is first used by Ludvig Lorenz (who invented retarded potential) but there is a more famous Hendrik Lorentz (the Lorentz transformation, Lorentz force, ...).

a wave equation (hence we have the name gravitational "waves")

$$\Box\bar{h}_{\mu\nu} = -16\pi G T_{\mu\nu}. \tag{1.18}$$

There are naturally ten degrees of freedom, but the Lorenz gauge condition imposes four constraints, leaving only six degrees of freedom. Some of these can still be fixed by applying additional gauges and we will leave this to the next section. The Lorenz gauge also implies

$$\partial^{\nu} T_{\mu\nu} = 0, \tag{1.19}$$

which is the conservation of energy-momentum tensor.

The calculations involving metric perturbations and approximations to linear order, as shown above, are referred to as *linearized theory*. Linearized gravity is effective when the gravitational field is weak, which is often the case for gravitational waves. For perturbations on a curved background, we will face the problem of how to separate the background and the waves. It is hard to tell the fundamental distinction between a curved spacetime and small fluctuations over it, but on a macroscopic level, the spacetime curvature and fluctuations may have different length scales and therefore be separable. This is important when it comes to the energy carried by GWs, and there has been a debate about whether GWs can carry energy, or, whether there is a well-defined energy-momentum tensor for GWs. The answer is yes, and details can be found in, e.g., Refs. [5, 6]

1.1.2 Polarizations of gravitational waves

Let us consider the propagation of GWs in the vacuum. The right-hand side of Eq. 1.18 vanishes, which leads to

$$\Box\bar{h}_{\mu\nu} = 0, \tag{1.20}$$

and the plane waves solution is simply

$$\bar{h}_{\mu\nu} = C_{\mu\nu} e^{ik_{\sigma}x^{\sigma}}, \tag{1.21}$$

where $k_{\sigma} = (\omega, k_1, k_2, k_3)$ is the 4-wavevector, x^{σ} is the position vector, and $C_{\mu\nu}$ is the (symmetric) polarization tensor of GWs. Substituting the solution Eq. 1.21 back to wave equation Eq. 1.20, we get

$$k^{\mu}k_{\mu} = 0, \tag{1.22}$$

which means GWs travel along null trajectories, in other words, at the speed of light. Substituting the solution Eq. 1.21 into Lorenz gauge Eq.1.10, we get

$$k^{\mu}C_{\mu\nu} = 0, \tag{1.23}$$

which means the wavevector is orthogonal to the polarization tensor, eliminating four degrees of freedom of $C_{\mu\nu}$. This leaves us with six independent components of $C_{\mu\nu}$, corresponding to the number of independent equations in Eq. 1.20. However, the Lorenz gauge $\partial^\mu \bar{h}_{\mu\nu} = 0$ does not fully fix all degrees of freedom: we can still perform coordinate transformation $x^\mu \to x'^\mu = x^\mu + \zeta^\mu$ without breaking Lorenz gauge as long as ζ^μ satisfies $\Box\zeta^\mu = 0$. There are four additional degrees of freedom inside ζ^μ to be fixed, so finally there should only be two degrees of freedom. We choose ζ^μ with the following procedure:

1. The first component ζ^0 can be chosen such that the trace $\bar{h} = 0$. We get $\bar{h}_{\mu\nu} = h_{\mu\nu}$ after doing this and we can omit the bar over $h_{\mu\nu}$ thereafter.

2. The three spatial components ζ^i can be chosen such that $h^{0i} = 0$. Substituting this into Lorenz gauge we get $\partial_0 h^{00} = 0$, which means h^{00} is a constant and can be set to 0 with a trivial coordinate transformation too. Putting these together we get $h^{0\mu} = 0$.

In fact, ζ^μ can be constructed to satisfy both the wave equation and the aforementioned gauge requirements:

$$\begin{cases} \zeta_\mu = B_\mu e^{ik_\sigma x^\sigma} \\ B_0 = -\frac{i}{2k_0}\left(C_{00} + \frac{1}{2}C^\mu_\mu\right) \\ B_j = \frac{i}{2k_0^2}\left[-2k_0 C_{0j} + k_j\left(C_{00} + \frac{1}{2}C^\mu_\mu\right)\right] \end{cases} \tag{1.24}$$

In conclusion, the polarization tensor $C_{\mu\nu}$ (or the solution $h_{\mu\nu}$) should follow

$$C^{0\mu} = 0,\ C^i_i = 0,\ \partial^i C_{ij} = 0, \tag{1.25}$$

and these conditions collectively define the Transverse-Traceless gauge (TT gauge). There is no residual gauge freedom under the TT gauge, and GWs are characterized by only two independent polarizations. Consider GWs propagating along z direction and the plane wave solution Eq. 1.21, the metric perturbation can be expressed as

$$h^{\mathrm{TT}}_{\mu\nu}(t,z) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}_{\mu\nu} \cos[\omega(t-z)] = (h_+ C_+ + h_\times C_\times)_{\mu\nu}\cos[\omega(t-z)] \tag{1.26}$$

and

$$C_+ = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, C_\times = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{1.27}$$

are the two polarization tensors of GWs. They are called "plus" and "cross" polarizations (Fig. 1.1), in the sense that they respectively modify the proper distance (and hence the motion of test particles) in the "plus" and "cross" directions. Consider the interval ds^2

$$\begin{aligned} ds^2 =& -c^2dt^2 + \{1 + h_+ \cos[\omega(t-z)]\}\, dx^2 + \{1 - h_+ \cos[\omega(t-z)]\}\, dy^2 \\ &+ 2h_\times \cos[\omega(t-z)]dxdy + dz^2, \end{aligned} \tag{1.28}$$

and two points along x-axis at x_1 and x_2 ($x_2 - x_1 = L$). When a GW contains only plus polarization passes by, the proper distance between the two points is

$$s = (x_2 - x_1)\left[1 + h_+ \cos(\omega t)\right]^{1/2} \simeq L\left[1 + \frac{1}{2}h_+ \cos(\omega t)\right]. \tag{1.29}$$

Since C_+ only has two non-zero opposite components at xx and yy, the proper distance between x_1 and x_2 changes periodically along x and y direction: x-direction expands while y-diretion contracts, and then this alternation reverses, forming a "+" shape. Similarly, the effect of the cross mode is a rotation of 45° of the plus'. This effect can also be understood by observing a circle of stationary test masses on the $x - y$ plane. The separation of test masses S^μ is governed by the geodesic deviation equation

$$\frac{D^2}{D\tau^2}S^\mu = R^\mu{}_{\nu\rho\sigma}U^\nu U^\rho S^\sigma, \tag{1.30}$$

where $\frac{D}{D\tau}$ is the covariant derivative and U^ν is the four-velocity of test masses. To the leading order, corrections of U^ν by h can be ignored so it is simply $(1, 0, 0, 0)$, and we can use $\tau \sim t$ for the slowly-moving particles. The Riemann tensor reduces to $R_{\mu 00\sigma} = \frac{1}{2}\partial_0\partial_0 h_{\mu\sigma}$ in the equation, and the geodesic deviation equation takes the form

$$\ddot{S}^\mu = \frac{1}{2}\ddot{h}^{\mathrm{TT}}_{\mu\nu}S^\nu. \tag{1.31}$$

Consider $\delta x = S^1 - x_0$ and $\delta y = S^2 - y_0$, they can be easily solved when we separate plus and cross polarizations:

$$\begin{aligned} \delta x(t) &= \frac{h_+}{2}x_0 \sin\omega t,\ \ \delta y(t) = -\frac{h_+}{2}y_0 \sin\omega t \quad \text{(plus)}, \\ \delta x(t) &= \frac{h_\times}{2}y_0 \sin\omega t,\ \ \delta y(t) = \frac{h_\times}{2}x_0 \sin\omega t \quad \text{(cross)}. \end{aligned} \tag{1.32}$$

The motion of the test masses is visualized in Fig. 1.1. h_+ and $h_\times$ stretch and compress the ring in different directions, composing the shapes of "+" and "×".

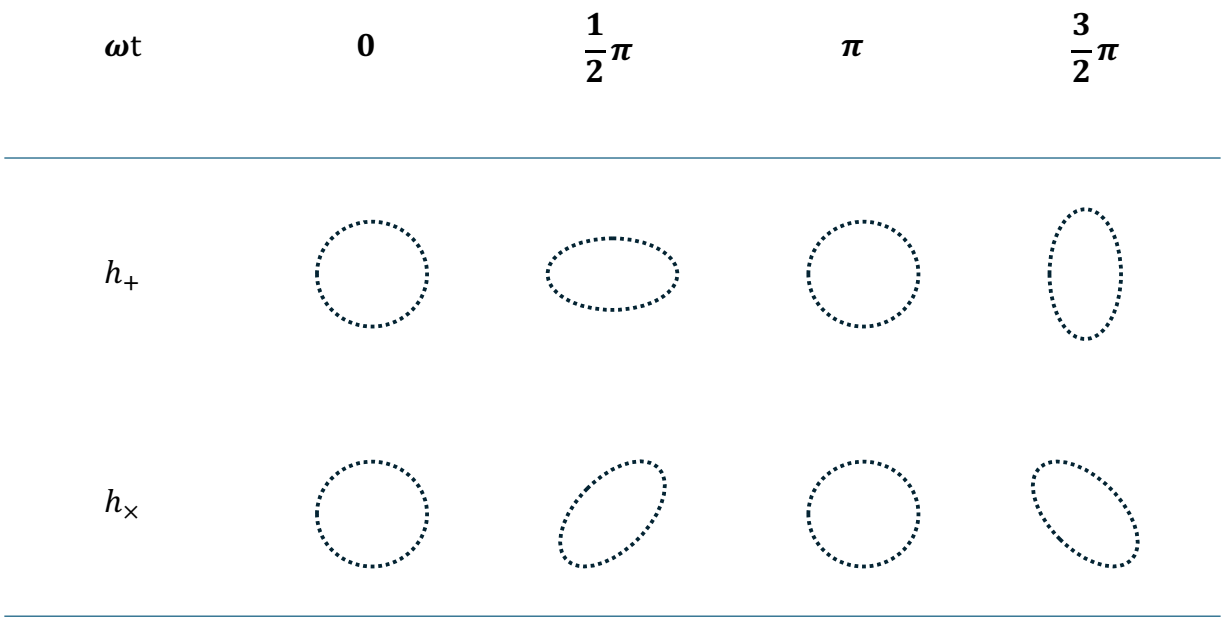


Figure 1.1: Illustrations of the effects of two polarizations of GWs. Test messes are arranged in a ring on the $x-y$ plane and the ring is stretched and compressed when the plus (middle row) and cross (bottom row) polarizations GW pass by along the z-axis.

1.1.3 Generation of gravitational waves

The plane-wave solution is a showcase of how GWs behave in the vacuum. However, to understand how gravitational waves are generated from physical sources, we need to solve Eq. 1.18. We can use the same method in solving Eq. 1.16 with the Green's function defined in Eq. 1.17. However, now we are considering waves that move forward in the time coordinate: the waves at position x^i at time x^0 are generated by the source at y^i at a time t^0 before x^0, therefore, we should use the retarded Greens function

$$G\left(x^\sigma - y^\sigma\right) = -\frac{1}{4\pi|\mathbf{x}-\mathbf{y}|}\delta\left[|\mathbf{x}-\mathbf{y}| - \left(x^0 - y^0\right)\right]\theta\left(x^0 - y^0\right), \tag{1.33}$$

where $\mathbf{x} = (x^1, x^2, x^3)$, $\mathbf{y} = (y^1, y^2, y^3)$ are spatial vectors, and $\theta\left(x^0 - y^0\right) = 1$ when $x^0 - y^0 > 0$ and $= 0$ otherwise. The GW solution is then written as

$$\begin{aligned}\bar{h}_{\mu\nu}\left(x^\sigma\right) &= -16\pi G\int G\left(x^\sigma - y^\sigma\right)T_{\mu\nu}\left(y^\sigma\right)d^4y \\ &= 4G\int\frac{T_{\mu\nu}(t-|\mathbf{x}-\mathbf{y}|,\mathbf{y})}{|\mathbf{x}-\mathbf{y}|}d^3y.\end{aligned} \tag{1.34}$$

Here $x^\sigma = (t, \mathbf{x})$. We integrated y^0 in the second line by limiting y^0 to $t - |\mathbf{x}-\mathbf{y}|$, i.e., the retarded time t_r. The physical picture of Eq. 1.34 is clear: the perturbation of gravitational field at $(t, \mathbf{x})$ arises from the cumulative influence from matter located at $\mathbf{y}$ at the retarded

time t_r.

We perform the Fourier transform to $\bar{h}_{\mu\nu}$. The Fourier transform and its inverse are defined as

$$\begin{aligned}\widetilde{\phi}(\omega,\mathbf{x}) &= \frac{1}{\sqrt{2\pi}}\int dt e^{i\omega t}\phi(t,\mathbf{x}),\\ \phi(t,\mathbf{x}) &= \frac{1}{\sqrt{2\pi}}\int d\omega e^{-i\omega t}\widetilde{\phi}(\omega,\mathbf{x}),\end{aligned} \tag{1.35}$$

and the Fourier transform of $\bar{h}_{\mu\nu}$ can be calculated by constructing a Fourier transform of the source $T_{\mu\nu}$:

$$\begin{aligned}\tilde{\bar{h}}_{\mu\nu}(\omega,\mathbf{x}) &= \frac{4G}{\sqrt{2\pi}}\int dt_r d^3y e^{i\omega t_r}e^{i\omega|\mathbf{x}-\mathbf{y}|}\frac{T_{\mu\nu}\left(t_r,\mathbf{y}\right)}{|\mathbf{x}-\mathbf{y}|}\\ &= 4G\int d^3y e^{i\omega|\mathbf{x}-\mathbf{y}|}\frac{\widetilde{T}_{\mu\nu}(\omega,\mathbf{y})}{|\mathbf{x}-\mathbf{y}|}.\end{aligned} \tag{1.36}$$

To further simplify this equation, we consider GWs from a distant isolated source and assume the source is moving at a low speed. With the first approximation, the source can be considered as at the distance $R \approx |\mathbf{x}-\mathbf{y}|$ and its size δR is much smaller than R. With the second, we can conclude $\omega\delta R \ll 1$, i.e. the motion of parts of the source is much slower than the speed of GWs (light). Putting these together, we can substitute $e^{i\omega|\mathbf{x}-\mathbf{y}|}/|\mathbf{x}-\mathbf{y}|$ with $e^{i\omega R}/R$, which can be extracted from the integral:

$$\tilde{\bar{h}}_{\mu\nu}(\omega,\mathbf{x}) = 4G\frac{e^{i\omega R}}{R}\int d^3y\widetilde{T}_{\mu\nu}(\omega,\mathbf{y}). \tag{1.37}$$

Therefore, we only need to calculate the integral of $\widetilde{T}_{\mu\nu}$ over spatial coordinates. This can be accomplished by integrating by parts and applying the conservation law Eq. 1.19. For example, $\widetilde{T}_{ij}$:

$$\begin{aligned}\int d^3y\widetilde{T}^{ij}(\omega,\mathbf{y}) &= \int\partial_k\left(y^i\widetilde{T}^{kj}\right)d^3y - \int y^i\left(\partial_k\widetilde{T}^{kj}\right)d^3y\\ &= i\omega\int y^i\left(\widetilde{T}^{0j}\right)d^3y\\ &= \frac{i\omega}{2}\int\left[\partial_l\left(y^iy^j\widetilde{T}^{0l}\right) - y^iy^j\left(\partial_l\widetilde{T}^{0l}\right)\right]d^3y\\ &= -\frac{\omega^2}{2}\int y^iy^j\widetilde{T}^{00}d^3y.\end{aligned} \tag{1.38}$$

Here we used integral by parts twice (the first and the third lines) and abandoned the boundary terms since the source is isolated. For the second and the fourth lines, we used the conservation law in the frequency domain:

$$\partial_\mu\widetilde{T}^{\mu\nu} = \partial_0\widetilde{T}^{0\nu} + \partial_i\widetilde{T}^{i\nu} = i\omega\widetilde{T}^{0\nu} + \partial_i\widetilde{T}^{i\nu} = 0. \tag{1.39}$$

Define the mass quadrupole moment tensor of the source

$$Q_{ij}(t) = \int y^i y^j T^{00}(t, \mathbf{y}) d^3 y, \tag{1.40}$$

the frequency domain solution can be written as

$$\tilde{\bar{h}}_{ij}(\omega, \mathbf{x}) = -2G\omega^2 \frac{e^{i\omega R}}{R} \widetilde{Q}_{ij}(\omega), \tag{1.41}$$

and we can convert it to the time domain

$$\begin{aligned} \bar{h}_{ij}(t, \mathbf{x}) &= -\frac{1}{\sqrt{2\pi}} \frac{2G}{R} \int d\omega e^{-i\omega(t-R)} \omega^2 \widetilde{Q}_{ij}(\omega) \\ &= \frac{1}{\sqrt{2\pi}} \frac{2G}{R} \frac{d^2}{dt^2} \int d\omega e^{-i\omega t_r} \widetilde{Q}_{ij}(\omega) \\ &= \frac{2G}{R} \ddot{Q}_{ij}\left(t_r\right), \end{aligned} \tag{1.42}$$

Eq. 1.42 is the quadrupole radiation formula. The time components of $\bar{h}_{\mu\nu}$ are given by the Lorenz gauge condition Eq. 1.10

$$\tilde{\bar{h}}^{0\nu} = \frac{i}{\omega} \partial_i \tilde{\bar{h}}^{i\nu}. \tag{1.43}$$

Note that time components of $\bar{h}_{\mu\nu}$ vanish in TT gauge. Tensors that are not under TT gauge can be transformed to TT gauge by

$$\bar{h}_{ij}^{\mathrm{TT}} = \Lambda_{ij,k\ell} \bar{h}_{k\ell}, \tag{1.44}$$

where the lambda projector is defined as

$$\Lambda_{ij,k\ell}(\hat{\mathbf{n}}) = P_{ik} P_{j\ell} - \frac{1}{2} P_{ij} P_{k\ell}, \qquad P_{ij}(\hat{\mathbf{n}}) = \delta_{ij} - n_i n_j, \tag{1.45}$$

where n_i is the direction of GW propagation. Therefore, spatial components under TT gauge should be written as

$$\bar{h}_{ij}^{\mathrm{TT}}(t, \mathbf{x}) = \frac{2G}{R} \ddot{\Lambda}_{ij,k\ell} Q_{kl}\left(t_r\right) = \frac{2G}{R} \ddot{Q}_{ij}^{\mathrm{TT}}\left(t_r\right). \tag{1.46}$$

This tells us that, the leading term of gravitational radiation is quadrupole radiation, which is proportional to the second derivative of the mass quadrupole momentum. Gravitational radiation under GR does not contain dipole radiation as electromagnetic radiation does. This is because changes in the dipole momentum of mass correspond to the motion of the center of mass, and in GR we can always find a set of coordinates to remove its

motion.

The quadrupole momentum is typically much smaller than the dipole since it is a higher-order description of the shape of the system, and due to the weak nature of gravity, we can infer that gravitational radiation is correspondingly very weak. We could also see this by writing Eq. 1.46 in the International System of Units (SI): the coefficient $2G/R$ will become $2G/Rc^4$. For a GW source at 1Mpc, it is a factor of 5×10^{-67}! This makes the detection of GWs extremely challenging, leading Einstein to conclude that although GWs existed, they were unlikely to ever be detected. However, Einstein underestimated the rapid advancement of technology and the determination of scientists. In 2015, the Laser Interferometer Gravitational-Wave Observatory (LIGO) has made the first direct detection of GWs from a Binary Black Hole (BBH) coalescence [7]. Further details will be provided in subsequent sections.

We are also interested in the angular distribution of GWs, since we have to observe GW sources from a specific angle. Let us set up the geometry as in Fig. 1.2: there is a Cartesian coordinate (x, y, z) in the source frame, and the observer's line of sight is along vector $\hat{\mathbf{n}}$. The angle between $\hat{\mathbf{n}}$ and $\hat{z}$ is θ. We construct another Cartesian coordinate (x', y', z') for the observer, where $\hat{\mathbf{z}'} = \hat{\mathbf{n}}$, and $\hat{\mathbf{z}'}$'s projection on (x, y) plane forms an angle of ϕ with respect to y. We aim to calculate h_+^{TT} and $h_\times^{\mathrm{TT}}$ for the observer, i.e., in the observer frame, via Eq. 1.46.

The first step would be finding the mass quadrupole in the TT gauge. Notice that when $\hat{\mathbf{n}} = \hat{\mathbf{z}'}$, the P_{ij} in the Lambda projector Eq. 1.45 is a diagonal matrix $\mathrm{diag}(1, 1, 0)$, therefore in the (x', y', z') frame, the derivative of mass quadrupole Q'_{ij} after TT projection can be calculated:

$$\ddot{Q}'^{\mathrm{TT}}_{ij} = \Lambda_{ij,kl}\ddot{Q}'_{kl} = \begin{pmatrix} \left(\ddot{Q}'_{11} - \ddot{Q}'_{22}\right)/2 & \ddot{Q}'_{12} & 0 \\ \ddot{Q}'_{21} & -\left(\ddot{Q}'_{11} - \ddot{Q}'_{22}\right)/2 & 0 \\ 0 & 0 & 0 \end{pmatrix}_{ij}. \tag{1.47}$$

Then the plus and cross polarizations for observers at $\hat{\mathbf{n}}$ direction are simply

$$\begin{aligned} h_+(t, \hat{\mathbf{n}}) &= \frac{G}{r}\left(\ddot{Q}'_{11} - \ddot{Q}'_{22}\right) \\ h_\times(t, \hat{\mathbf{n}}) &= \frac{2G}{r}\ddot{Q}'_{12}, \end{aligned} \tag{1.48}$$

Finally, from the coordinate transformation, Q'_{ij} is related to the source frame mass quadrupole Q_{ij} by

$$Q_{ij} = \mathcal{R}_{ik}Q'_{kl}\mathcal{R}_{jl}, \tag{1.49}$$

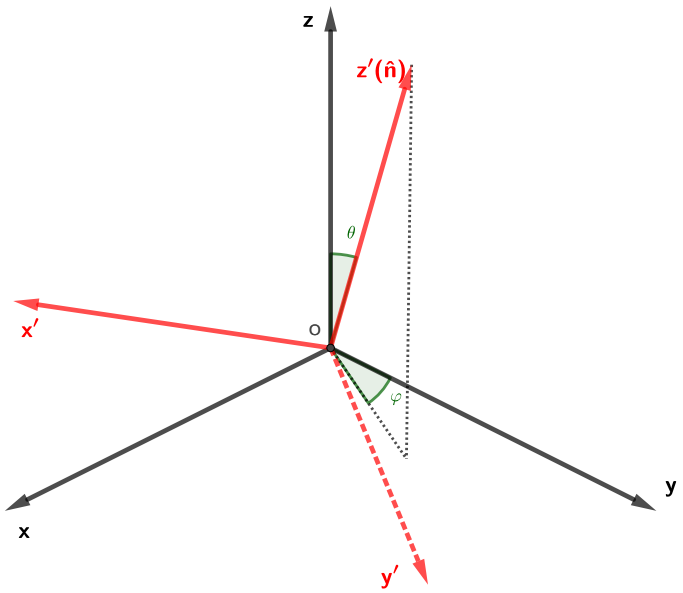


Figure 1.2: The relation between the source frame (x, y, z) and the observer frame (x', y', z'). The observer's line of sight is along z', and the angle between $\hat{\mathbf{n}}$ and $\hat{z}$ is θ. $\hat{\mathbf{z}}'$'s projection on (x, y) plane has an angle of ϕ with respect to y.

where

$$\mathcal{R} = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix}. \tag{1.50}$$

To summarize, one can calculate the mass quadrupole in the source frame without concerning the TT gauge, and convert it to the observer frame by Eq. 1.49, and then the GW polarizations observed are given by Eq. 1.48.

1.1.4 Energy carried by gravitational waves

GWs carry energy, yet defining the energy of GWs poses several conceptual challenges [5]. First, the energy of gravitational fields can not be defined locally as electromagnetic fields can. Electromagnetic fields are essentially matter that "has weight"; they curve the space-time and leave observable signatures. This is not the case for gravitational fields. The equivalence principle tells us that the local curvature can always be eliminated by coordinate transformations. Without a gravitational field, there is no well-defined local gravitational energy. Therefore, gravitational energy, including GWs' energy, should always be defined on the macroscopic scale. Second, when considering GWs' energy (i.e.

their energy-momentum tensors), we must include it in the Einstein field equations and it will in turn curve its background spacetime. In earlier sections, we perturbed a flat background spacetime to distinguish GWs. Now, with the background curved, it is necessary to ensure that the wavelength $\lambda_{\rm GW}$ of the GW is much smaller than the spatial variation of the background gravitational field $L_{\rm BG}$, i.e. $\lambda_{\rm GW} \ll L_{\rm BG}$. Only under these conditions does it become meaningful to discuss the energy of GWs.

The Einstein field equations can be rearranged as

$$R_{\mu\nu} = 8\pi G \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T \right), \tag{1.51}$$

and we expand the Ricci tensor $R_{\mu\nu}$ to $\mathcal{O}(h^2)$:

$$R_{\mu\nu} = \bar{R}_{\mu\nu} + R^{(1)}_{\mu\nu} + R^{(2)}_{\mu\nu} + \dots, \tag{1.52}$$

where $\bar{R}_{\mu\nu}$ is caused by the background $\bar{g}_{\mu\nu}$ only, $R^{(1)}_{\mu\nu}$ is of $\mathcal{O}(h)$ and $R^{(2)}_{\mu\nu}$ is of $\mathcal{O}(h^2)$. Their explicit expressions are [6]

$$R^{(1)}_{\mu\nu} = \frac{1}{2} \left(\bar{\nabla}^\alpha \bar{\nabla}_\mu h_{\nu\alpha} + \bar{\nabla}^\alpha \bar{\nabla}_\nu h_{\mu\alpha} - \bar{\nabla}^\alpha \bar{\nabla}_\alpha h_{\mu\nu} - \bar{\nabla}_\nu \bar{\nabla}_\mu h \right), \tag{1.53}$$

$$\begin{aligned} R^{(2)}_{\mu\nu} =& \frac{1}{2} \bar{g}^{\rho\sigma} \bar{g}^{\alpha\beta} \left[\frac{1}{2} \bar{\nabla}_\mu h_{\rho\alpha} \bar{\nabla}_\nu h_{\sigma\beta} + \left(\bar{\nabla}_\rho h_{\nu\alpha} \right) \left(\bar{\nabla}_\sigma h_{\mu\beta} - \bar{\nabla}_\beta h_{\mu\sigma} \right) \right. \\ & + h_{\rho\alpha} \left(\bar{\nabla}_\nu \bar{\nabla}_\mu h_{\sigma\beta} + \bar{\nabla}_\beta \bar{\nabla}_\sigma h_{\mu\nu} - \bar{\nabla}_\beta \bar{\nabla}_\nu h_{\mu\sigma} - \bar{\nabla}_\beta \bar{\nabla}_\mu h_{\nu\sigma} \right) \\ & \left. + \left(\frac{1}{2} \bar{\nabla}_\alpha h_{\rho\sigma} - \bar{\nabla}_\rho h_{\alpha\sigma} \right) \left(\bar{\nabla}_\nu h_{\mu\beta} + \bar{\nabla}_\mu h_{\nu\beta} - \bar{\nabla}_\beta h_{\mu\nu} \right) \right], \end{aligned} \tag{1.54}$$

where $\bar{\nabla}_\mu$ is the covariant derivative with respect to the background metric. As discussed earlier, the background should be smooth and correspond to "low frequency" (slowly varying) terms in the equations, so $\bar{R}_{\mu\nu}$ only contains low frequency modes. Similarly, $R^{(1)}_{\mu\nu}$ only contains high frequency (fluctuating) modes. Interestingly, $R^{(2)}_{\mu\nu}$ should include both low and high frequency modes since GWs with different wave vectors can destructively interfere and result in a smooth perturbation. Separating low and high frequency modes of $R^{(2)}_{\mu\nu}$, we can write

$$\bar{R}_{\mu\nu} = -\left[R^{(2)}_{\mu\nu} \right]^{\rm Low} + 8\pi G \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T \right)^{\rm Low}, \tag{1.55}$$

$$R^{(1)}_{\mu\nu} = -\left[R^{(2)}_{\mu\nu} \right]^{\rm High} + 8\pi G \left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T \right)^{\rm High}. \tag{1.56}$$

How can we naturally divide the low and high frequency parts? Notice that the spatial scale $\bar{l}$ that is much greater than $\lambda_{\rm GW}$ but smaller than the background $L_{\rm BG}$, and if we

perform a spatial average over scale $\bar{l}$, the modes that are around $\lambda_{\rm GW}$ will be averaged out and only low frequency modes could remain. Therefore, Eq. 1.55 can also be written as

$$\bar{R}_{\mu\nu} = -\left\langle R^{(2)}_{\mu\nu} \right\rangle + 8\pi G \left\langle T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T \right\rangle . \tag{1.57}$$

where $\langle \dots \rangle$ means spatial average over several $\lambda_{\rm GW}$. We can further define the "averaged", low frequency part of $T_{\mu\nu}$, which is the effective energy-momentum tensor of the matter

$$\left\langle T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T \right\rangle = \bar{T}^{\mu\nu} - \frac{1}{2} \bar{g}_{\mu\nu} \bar{T}, \tag{1.58}$$

and we find if we let

$$t_{\mu\nu} = -\frac{1}{8\pi G} \left\langle R^{(2)}_{\mu\nu} - \frac{1}{2} \bar{g}_{\mu\nu} R^{(2)} \right\rangle , \tag{1.59}$$

then Eq. 1.57 can be written as

$$\bar{R}_{\mu\nu} - \frac{1}{2} \bar{g}_{\mu\nu} \bar{R} = 8\pi G \left(\bar{T}_{\mu\nu} + t_{\mu\nu} \right) , \tag{1.60}$$

just like the background spacetime is curved by matter $T_{\mu\nu}$, and another term Eq. 1.59 which is from GWs! Eq. 1.59 therefore becomes the definition of the energy-momentum tensor of GWs. With the definition of $\bar{R}_{\mu\nu}$, one can verify that the conservation law $\bar{\nabla}^{\mu}(\bar{T}_{\mu\nu} + t_{\mu\nu}) = 0$, and this reduces to $\partial^{\mu} t_{\mu\nu} = 0$ in flat background far from the GW source. Substituting Eq. 1.54 into Eq. 1.59, $t_{\mu\nu}$ will be simplified as

$$t_{\mu\nu} = \frac{1}{32\pi G} \left\langle \partial_{\mu} h_{\alpha\beta} \partial_{\nu} h^{\alpha\beta} \right\rangle , \tag{1.61}$$

and we have

$$t^{00} = \frac{1}{32\pi G} \left\langle \dot{h}^{\rm TT}_{ij} \dot{h}^{\rm TT}_{ij} \right\rangle = \frac{1}{16\pi G} \left\langle \dot{h}^2_{+} + \dot{h}^2_{\times} \right\rangle \tag{1.62}$$

under the TT gauge. The energy flux on area element dA of GWs takes the form

$$\frac{dE}{dAdt} = t^{00} = \frac{1}{32\pi G} \left\langle \dot{h}^{\rm TT}_{ij} \dot{h}^{\rm TT}_{ij} \right\rangle , \tag{1.63}$$

and it can also be written in terms of the solid angle $d\Omega = dA/r^2$

$$\frac{dE}{dt} = \frac{r^2}{32\pi G} \int d\Omega \left\langle \dot{h}^{\rm TT}_{ij} \dot{h}^{\rm TT}_{ij} \right\rangle . \tag{1.64}$$

Similarly, we can write down the momentum that GWs carry

$$\frac{dP^k}{dAdt} = t^{0k} = \left\langle \dot{h}^{\rm TT}_{ij} \partial^k h^{\rm TT}_{ij} \right\rangle , \tag{1.65}$$

$$\frac{dP^k}{dt} = -\frac{1}{32\pi G} r^2 \int d\Omega \left\langle \dot{h}_{ij}^{\mathrm{TT}} \partial^k h_{ij}^{\mathrm{TT}} \right\rangle . \tag{1.66}$$

The fact that GWs carry energy implies that the source of GWs will lose energy, which in turn affects the dynamical evolution of the source. For instance, in a binary system, this energy loss results in a gradual decrease in orbital radius, potentially leading to the eventual merger of the binary system, as observed in the detections by LIGO. In fact, even before the first direct detection by LIGO in 2015, the energy carried by GWs had been indirectly confirmed by Hulse and Taylor as early as 1974 [8]: they found the orbit decay in a binary pulsar system follows the prediction by GR.

It is noteworthy that the averaging technique used here is essentially the renormalization group transformation, which is important in the Quantum Field Theory (QFT). We integrated out the small fluctuations that happen on a small length scale (λ_{GW}) to get an effective theory on the length scale we are interested in ($\bar{l}$). The difference between renormalization here and in QFT is that here we integrated over the coordinate space to obtain a theory useful in a typical length scale, while in QFT we may consider the momentum space and cut high momentums (high energy) off to obtain a low-energy effective theory.

1.1.5 Gravitational waves from binary systems

We now turn our attention to a fundamental physical model: two point masses orbiting each other. This is one of the simplest models that has changes in mass quadrupole so it emits GWs. As binary systems are common in the universe [9, 10], studying this basic model will provide insights into astrophysical sources discussed in the following sections. We will calculate the leading-term GW waveform from it and investigate how such systems evolve.

We consider the geometry in Fig. 1.2 and assume the two objects m_1, m_2 and their circular orbit are on the (x, y) plane of the source frame. We set the origin of the source coordinates as the center of mass of the system and assume the distance between m_1 and m_2 is a. When they orbit around each other at the angular velocity ω_s, the mass density of the system is

$$\begin{aligned} T^{00}(t, \boldsymbol{x}) = \rho(t, \boldsymbol{x}) =& m_1 \left[\delta \left(x - r_1 \cos(\omega_s t + \frac{\pi}{2}) \right) \delta \left(y - r_1 \sin(\omega_s t + \frac{\pi}{2}) \right) \delta(z) \right] \\ &+ m_2 \left[\delta \left(x + r_2 \cos(\omega_s t + \frac{\pi}{2}) \right) \delta \left(y + r_2 \sin(\omega_s t + \frac{\pi}{2}) \right) \delta(z) \right], \end{aligned} \tag{1.67}$$

where

$$r_1 = \frac{m_2}{m_1 + m_2} a, \ r_2 = \frac{m_1}{m_1 + m_2} a. \tag{1.68}$$

The $\pi/2$ in the phase is chosen to simplify the result in the end (it removes an annoying minus sign). Substituting the mass density into Eq. 1.40, the non-zero mass quadrupole

components are

$$\begin{aligned} Q_{11} &= \mu a^2 \frac{1-\cos 2\omega_s t}{2} \\ Q_{22} &= \mu a^2 \frac{1+\cos 2\omega_s t}{2} \\ Q_{12} &= -\frac{1}{2}\mu a^2 \sin 2\omega_s t, \end{aligned} \tag{1.69}$$

where $\mu = m_1 m_2/(m_1+m_2)$ is the reduced mass. The derivatives are

$$\begin{aligned} \ddot{Q}_{11} &= -\ddot{Q}_{22} = 2\mu a^2 \omega_s^2 \cos 2\omega_s t, \\ \ddot{Q}_{12} &= 2\mu a^2 \omega_s^2 \sin 2\omega_s t. \end{aligned} \tag{1.70}$$

Combining the mass quadrupole with Eq. 1.48 and Eq. 1.49, we obtain the GW waveform

$$\begin{aligned} h_+(t;\theta,\phi) &= \frac{4G\mu\omega_s^2 a^2}{r}\left(\frac{1+\cos^2\theta}{2}\right)\cos\left(2\omega_s t_{\rm ret} + 2\phi\right) \\ h_\times(t;\theta,\phi) &= \frac{4G\mu\omega_s^2 a^2}{r}\cos\theta \sin\left(2\omega_s t_{\rm ret} \ + 2\phi\right). \end{aligned} \tag{1.71}$$

The frequency of the quadrupole gravitational radiation is twice the source orbit frequency. The ϕ only contributes a phase to the waveform; it degenerates with a time translation of the binary, i.e., one can shift the system by Δt and replace ϕ with another angle without changing the physics. Thus, we can use the observer time t to replace the source fame retarded time:

$$2\omega_s t_{\rm ret} \ + 2\phi \to 2\pi f t + \phi_0 \tag{1.72}$$

An interesting observation is that the amplitude of GWs depends on θ, which in this context represents the *inclination angle* ι of the source. ι describes the rotational axis of the system and the line of sight of the observer. We always observe the plus polarization, but when the source is "edge on" toward us ($\iota = \pi/2$), the cross polarization vanishes. When we are right above the binary plane, the "face-on" sources ($\iota = 0$) have the largest amplitude.

The amplitude also depends on the mass of the system. However, it is not directly proportional to μ as shown in Eq. 1.71 because mass and orbital parameters (ω_s, a) are not independent. They are related by the physics of the Newtonian circular orbit (to the lowest order):

$$\omega_s^2 = \frac{Gm}{(a/2)^3}, \tag{1.73}$$

where $m = m_1 + m_2$ is the total mass. We want to remove a redundant parameter in the amplitude - and it is better to be a here since a is physically less interesting than the other

two. Plugging $a = \sqrt[3]{\omega_s^2/Gm}$ into Eq. 1.71 and using Eq. 1.72, we get

$$\begin{aligned} h_+(t) &= \mathcal{A}\left(\frac{1+\cos^2\iota}{2}\right)\cos\left(2\pi f t + 2\phi_0\right) \\ h_\times(t) &= \mathcal{A}\cos\iota\sin\left(2\pi f t + 2\phi_0\right). \end{aligned} \tag{1.74}$$

where

$$\mathcal{A} = \frac{4}{r}(G\mathcal{M})^{5/3}(\pi f)^{2/3}, \tag{1.75}$$

and

$$\mathcal{M} = \mu^{3/5}m^{2/5} = \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}} \tag{1.76}$$

is the *chirp mass* of the system. It is the chirp mass that directly determines the amplitude of GWs of binary systems (in the leading order). The power of the gravitational radiation can be calculated by integrating over the solid angle from Eq. 1.64, and it is also determined by the chirp mass:

$$P = \frac{r^2}{16\pi G}\int d\Omega\left\langle \dot{h}_+^2 + \dot{h}_\times^2\right\rangle = \frac{32}{5G}\left(\pi G\mathcal{M}f\right)^{10/3}. \tag{1.77}$$

The power of gravitational radiation brings a new problem: the binary cannot maintain a stable circular orbit due to the loss of energy by GWs. We have to consider the back-reaction of GWs on the system. To the Newtonian order, the energy of the system is

$$E = -\frac{Gm_1m_2}{a} = -\left(\frac{\pi^2G^2\mathcal{M}^5f^2}{8}\right)^{1/3}, \tag{1.78}$$

note here f is the frequency of GWs. The decrease of energy will therefore cause the shrinking of orbit size a and increase of GW frequency f (as well as the orbit frequency ω_s). Consequently, on a sufficiently long timescale, the energy loss will lead to the *coalescence* of the binary. To provide a clearer picture, we assume the change in orbit is much slower than the velocity of the system, i.e., $|\dot{a}| \ll \omega_s a$. This allows us to treat the orbit as circular but with a slowly-varying radius. We are in the *quasi-circular* regime as long as this approximation holds.

We consider the loss of energy to be totally caused by GWs, then we can equate the time derive of Eq. 1.78 with Eq. 1.77, and get a differential equation for GW frequency f:

$$\dot{f} = \frac{96}{5}\pi^{8/3}(G\mathcal{M})^{5/3}f^{11/3}, \tag{1.79}$$

whose solution reads

$$f(\tau) = \frac{1}{8\pi}\left(\frac{5}{\tau}\right)^{3/8}(G\mathcal{M})^{-5/8}, \tag{1.80}$$

where $\tau = t_c - t$ is the time to coalescence and t_c is the time of coalescence. It is useful to write it and its inverse in SI:

$$\begin{aligned} f(\tau) &= \frac{1}{8\pi}\left(\frac{5}{\tau}\right)^{3/8}\left(\frac{G\mathcal{M}}{c^3}\right)^{-5/8} \\ &\approx 134 \text{ Hz}\left(\frac{1.21M_\odot}{\mathcal{M}}\right)^{5/8}\left(\frac{1s}{\tau}\right)^{3/8}, \end{aligned} \tag{1.81}$$

$$\begin{aligned} \tau(f) &= \frac{5}{256}\left(\frac{G\mathcal{M}}{c^3}\right)^{-5/3}(\pi f)^{-8/3} \\ &\approx 2.18 \text{ s}\left(\frac{1.21M_\odot}{\mathcal{M}}\right)^{5/3}\left(\frac{100 \text{ Hz}}{f_{\rm gw}}\right)^{8/3}. \end{aligned} \tag{1.82}$$

Eq. 1.80 tells us although the GW frequency evolves, there is a clear one-to-one relationship between the frequency and time to coalescence. Therefore, in the waveform formula, we can change f with $f(\tau)$, and replace $2\pi ft$ with a cumulative phase:

$$\Phi(t) = \int_{t_0}^{t} 2\pi f(t')dt'. \tag{1.83}$$

Finally, the waveform reads

$$\begin{aligned} h_+(t) &= \mathcal{A}(\tau)\left(\frac{1+\cos^2\iota}{2}\right)\cos[\Phi(\tau)+\phi_c] \\ h_\times(t) &= \mathcal{A}(\tau)\cos\iota\sin[\Phi(\tau)+\phi_c] \end{aligned} \tag{1.84}$$

where

$$\mathcal{A}(\tau) = \frac{1}{r}(G\mathcal{M})^{5/4}\left(\frac{5}{\tau}\right)^{1/4}, \tag{1.85}$$

$$\Phi(\tau) = -2\left(\frac{\tau}{5G\mathcal{M}}\right)^{5/8}. \tag{1.86}$$

Here Φ_c is an integral constant denoting the coalescence phase. Now we have quantified the evolution of the GW frequency and amplitude. As both of them increase and increase faster with time, the resulting signal resembles a chirp, and therefore we name the dominant parameter $\mathcal{M}$ as chirp mass. It's important to note that this characterization holds precisely for quasi-circular orbits, where the orbit remains circular and only the radius changes. In contrast, for eccentric orbits, GW frequency spectrum is not quasi-monochromatic and will extend to higher frequencies [6], which could render the signals sound less "chirpy" even though the overall frequency is increasing. Eccentric binaries lose energy faster than the quasi-circular, so their orbit evolution is faster than described in Eq. 1.82. Their eccentricity also decreases in this process, thus the orbit will be circularized during evolution. More details of eccentric binaries can be found in Ref. [5, 6], and

we will focus on quasi-circular binaries in this work.

The GW waveform in the frequency domain can be obtained by Fourier transform and Stationary Phase Approximation (SPA). SPA is based on the Riemann-Lebesgue lemma: for an integrable function $f(t)$ in $[a, b]$, we have

$$\lim_{\lambda \to \pm\infty} \int_a^b f(t) e^{i\lambda t} dt = 0. \tag{1.87}$$

An intuitional explanation to the lemma is that $e^{i\lambda t}$ contributes a fast oscillating term, and the integral is averaged out to zero due to the oscillations. On the contrary, if $e^{i\lambda t}$ does not oscillate in some places, these places will contribute most to the integral. For example, if the exponent λt, or a more generic form $\Phi(t)$, as a stationary point $\dot{\Phi} = 0$, the integral will be mostly contributed from the stationary point.

This is exactly the case when we calculate the Fourier transform of the time domain waveform Eq. 1.84. Take plus polarization for example:

$$\begin{aligned} \tilde{h}_+(f) &= \int dt \mathcal{A}\left(t_{\rm ret}\right) \cos\Phi\left(t_{\rm ret}\right) e^{i2\pi f t} \\ &= \frac{1}{2} e^{i2\pi f r} \int dt_{\rm ret} \mathcal{A}\left(t_{\rm ret}\right) \left(e^{i\Phi(t_{\rm ret})} + e^{-i\Phi(t_{\rm ret})}\right) e^{i2\pi f t_{\rm ret}}, \end{aligned} \tag{1.88}$$

where we have used $\tau = t_c - t$ and $t_{\rm ret} = t - r$. Since $\dot{\Phi(t)} > 0$, the $e^{i\Phi(t_{\rm ret})}$ will be always fast oscillating and it vanishes after integral[2]. The $e^{-i\Phi(t_{\rm ret})}$ together with $e^{i2\pi f t_{\rm ret}}$ will contribute a stationary point t_* satisfying

$$\dot{\Phi}(t_*) = 2\pi f, \tag{1.89}$$

so we can expand the exponential term to the second order around t_* and approximate the integral with the expanded phase. From a physics perspective, calculating $\tilde{h}_+(f)$ by only taking contributions around t_* means we only include frequencies near $f(t_*)$, aligning with our quasi-circular approximation. The expanded exponential term is

$$-2\Phi(t) + 2\pi f t = -2\Phi\left(t_*\right) + 2\pi f t_* - 2\pi \dot{f}\left(t_*\right)\left(t - t_*\right)^2 + \cdots, \tag{1.90}$$

and the Fourier transform becomes analytically integrable, which gives the result

$$\tilde{h}_+(f) = \frac{1}{2} e^{i\Psi_+} A\left(t_*\right) \left(\frac{2\pi}{\ddot{\Phi}\left(t_*\right)}\right)^{1/2} \tag{1.91}$$

where

$$\Psi_+ = 2\pi f\left(t_* + r\right) - \Phi\left(t_*\right) - (\pi/4). \tag{1.92}$$

[2]We only consider positive frequency here, and the negative frequency can be given by $\tilde{h}_+(-f) = \tilde{h}^*(f)$

The t_* can be converted to f via Eq. 1.80, and the final results read

$$\begin{aligned}\tilde{h}_+(f) &= \mathcal{A}(f)\left(\frac{1+\cos^2\iota}{2}\right)e^{i\Psi_+(f)}\\ \tilde{h}_\times(f) &= \mathcal{A}(f)\cos\iota e^{i\Psi_\times(f)}\end{aligned} \tag{1.93}$$

where

$$\mathcal{A}(f) = \frac{1}{\pi^{2/3}}\left(\frac{5}{24}\right)^{1/2}\frac{1}{r}\left(G\mathcal{M}\right)^{5/6}\frac{1}{f^{7/6}}, \tag{1.94}$$

$$\Psi_+(f) = 2\pi f\left(t_c + r\right) - \Phi_0 - \frac{\pi}{4} + \frac{3}{4}\left(8\pi G\mathcal{M}f\right)^{-5/3}, \tag{1.95}$$

and $\Psi_\times(f) = \Psi_+(f) + \pi/2$. Writing the amplitude in convenient units, we get an estimate of GW magnitude for binaries:

$$h \sim 1.4\times10^{-22}\left(\frac{f}{100\text{ Hz}}\right)^{2/3}\left(\frac{\mathcal{M}}{1.22M_\odot}\right)^{5/3}\left(\frac{10\text{Mpc}}{r}\right). \tag{1.96}$$

Note that $\mathcal{A}(\tau)$ diverges when $t \to t_c$. This is because when t approaches t_c the source has already lost the nice properties, such as components being far separated and at low speed, weak field etc, that allow us to calculate at the Newtonian order, therefore all of the above equations are not valid anymore. In fact, considering the Schwarzschild metric around the binary components, no stable circular orbit is allowed when they reach the Innermost Stable Circular Orbit (ISCO), the radius of which is given by the total mass m of the system [6]:

$$r_{\text{ISCO}} = 6Gm, \tag{1.97}$$

and it corresponds to the GW frequency of quasi-circular orbits

$$\begin{aligned}f^{\text{GW}}_{\text{ISCO}} &= \frac{1}{6\pi\sqrt{6}}\frac{1}{Gm}\\ &\approx 440\text{Hz}\left(\frac{10M_\odot}{m}\right)\quad\text{(in SI)}.\end{aligned} \tag{1.98}$$

After the ISCO is reached, the dynamics of the binary system will be driven by strong-field effects, ultimately leading to the merger of the two components. This will be discussed in more detail in later sections. Example waveforms before the ISCO are shown in Fig. 1.3. We can read the chirp characteristics from the waveforms, and how the inclination angles influence the amplitude of two polarizations.

At the end of this section, it is important to acknowledge that we only considered GW physics under GR. Despite GR's remarkable success and extensive experimental validation [11], it is facing difficulties on both theory and observation sides such as quantization, dark matter and dark energy. Many modified gravity theories have been proposed with

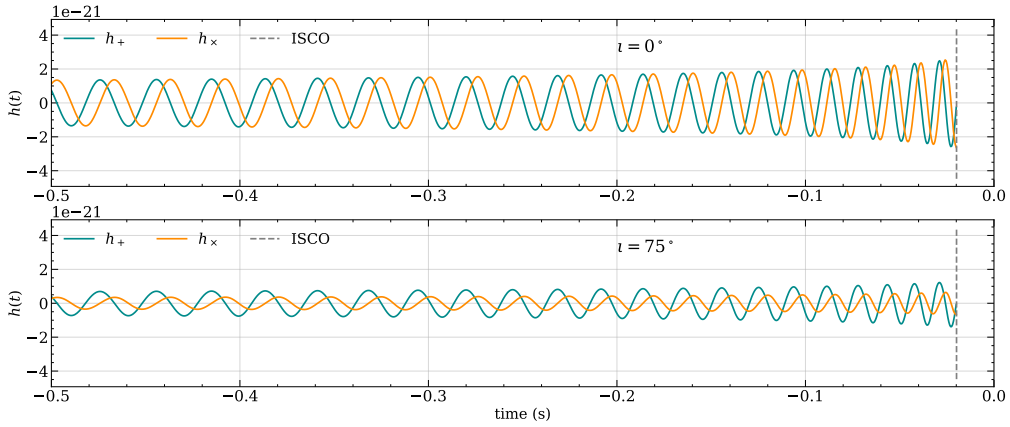


Figure 1.3: Example GW waveforms in time domain from a $20M_\odot$+$20M_\odot$ binary system at 100Mpc with different inclination angles (0 and 75°).

different motivations (e.g.,[12, 13, 14, 15, 16]), and GWs in those theories could behave differently from GWs in GR in generation (the source property has a different relation with GWs emitted), propagation (dispersion [17], birefringence [18], and amplitude damping [19, 20]), and polarization (up to 4 more extra polarizations of GWs [21, 22, 23, 24]). While it's beyond the scope of this thesis to provide an exhaustive review of these theories, it is essential to remain open to the possibility that GWs could behave differently in alternative theories of gravity. In fact, the observation of GWs can be used as a unique test of gravity theories [25].

1.2 Astrophysical sources of gravitational waves

GWs are extremely weak phenomena. The magnitude of GW sources can be roughly estimated by replacing the $\ddot{Q}_{ij}^{\rm TT}$ in Eq. 1.46 with Mv^2, where M is the mass and v is the typical velocity of the system. For a man-made source with two masses of $M = 10^3$kg orbiting each other with $v = 100$m/s, the h will be around 10^{-37} at $r = 1$m, which is far too small to be detected, not to mention the feasibility of making such a centrifuge system. Therefore, the hope for GW detection lies in those extremely energetic sources in the deep universe where masses can be several to billions of solar masses and velocities can approach the speed of light. Fig. 1.3 shows that a Binary Black Hole (BBH) system could produce $h \sim 10^{-21}$, which is much stronger than the man-made sources, suggesting the astrophysical sources are more promising. I will give an overview of astrophysical sources of GWs in this section.

1.2.1 Compact binaries

Binary systems are common in the universe, in fact, binary stars are more common than single stars [9, 10]. During the final stages of stellar evolution, massive stars may collapse into compact objects, including Black Holes (BHs), Neutron Stars (NSs), and White Dwarfs (WDs), depending on their mass. A binary star may go through a coevolution process and form a compact binary [9, 26, 27, 28]. Common stellar mass compact binaries include Binary Black Hole (BBH), Binary Neutron Star (BNS), Neutron Star Black Hole Binaries (NSBH), and Double White Dwarf (DWD). Besides, stellar mass compact binaries can form through the dynamical capture of single objects in dense star clusters or Active galactic nuclei (AGN) disks [29, 30, 31]. In rare cases, parabolic and hyperbolic encounters of two compact objects are possible, which could produce GWs analogous to Bremsstrahlung processes in electromagnetic radiation production [32, 33][3]. These compact binaries are natural massive centrifuges that generate "strong" GW signals and are ideal sources for GW detection.

In addition to the stellar-origin BHs and NSs, theories have predicted other alternatives that are yet to be confidently observed. For instance, primordial BH is an alternative that can be formed in the early universe [34, 35] and their merger rate is predicted in Ref. [36]. Exotic objects (like boson stars [37]) are also alternatives to BHs and NSs, and are predicted to be able to form in the universe [38, 39].

Beyond the stellar mass compact binaries above, it is also expected that massive black holes in the center of galaxies could also be sources of GWs when they interact with other heavy objects. This contains Super-Massive Binary Black Hole (SMBBH) [40, 41] (if the other object is also a supermassive black hole, which indicates the merger of two galaxies) and Extreme-Mass-Ratio Inspiral (EMRI) [42, 43] (if the other one is a stellar-mass object captured by the massive black hole).

The gravitational radiation and evolution of binary systems with quasi-circular orbits have been discussed in Sec. 1.1.5: Binary systems lose energy in the form of GWs during orbiting, which results in the decrease of the orbit radius and the increase of orbit frequency, thereby producing "chirp-like" GW signals. This is known as the *inspiral* stage, during which binary components are far separated and move at low velocities in a weak gravitational field. These properties break down when the components get so close to each other that the Schwarzschild metric no longer allows for stable circular orbits. After the ISCO is reached, the dynamics will be driven by strong-field effects and the binary will plunge into each other. This is referred to as the *merger* stage, during which the amplitude of GWs reaches its peak. The merger may lead to the formation of a larger BH [7] at an excited (perturbed) state, which will produce a damping GW signal due to Quasi-Normal

[3]It is not usually listed as a "compact binary" source in the literature. They are included here for completeness but will not be discussed in detail.

Modes (QNMs) of vibration [44]. The last stage is called *ringdown*. This whole inspiral-merger-ringdown process is often referred to as Compact Binary Coalescence (CBC).

However, depending on the type, mass, initial condition, and distance of the compact binaries, they will produce different kinds of GW signals, and not all compact objects can go through the aforementioned inspiral-merger-ringdown process. The first observation from Eq. 1.98 is that the frequency of GWs is inversely proportional to the mass of the source. For stellar-mass compact binaries, the frequency band lines in $\mathcal{O}(1) \sim \mathcal{O}(10^3)$Hz near merger, while for SMBBH and EMRI the frequency band is $\mathcal{O}(10^{-4}) \sim \mathcal{O}(10^{-1})$Hz, which implies different types of detectors may be needed to observe the coalescence stellar-mass and supermassive binaries. On the other hand, low-frequency detectors may detect the very early stage of stellar-mass binaries. This will be discussed in the next section about detectors.

Second, the time to merger formula Eq. 1.82 can be rearranged to estimate the time from formation to merger τ_0

$$\tau_0 \simeq 9.8 \times 10^6 \text{yr} \left(\frac{T_0}{1\text{hr}}\right)^{8/3} \left(\frac{M_\odot}{m}\right)^{2/3} \left(\frac{M_\odot}{\mu}\right), \tag{1.99}$$

where T_0 is the initial orbit period. Given the galaxy's age is estimated at 10 billion years [45], stellar mass binaries started close enough (small T_0) will merge at a timescale that we are able to observe. They will leave transient signals in high-frequency-sensitive detectors (which can only detect the late high-amplitude GWs), and leave long-lasting signals in low-frequency-sensitive detectors, and will be detected if the signal is strong enough. However, there is also a certain fraction of compact binaries that can not evolve to the merger stage within the current age of galaxies, or, they may not be observed when they merge because they are too distant from us. Since there are many of them, they will contribute a Stochastic Gravitational-Wave Background (SGWB) to the detector data, which can only be estimated at the population level statistically [46, 47].

In addition to the GW signals, some compact binaries are expected to have Electromagnetic (EM) counterparts as well as other counterparts like neutrinos, i.e., strong EM signals will be generated during the orbiting or merger. For example, the merger of BNS may be observed as a kilonova which is bright from GRB to radio band [48, 49]. A "wet" environment around the binary could also lead to EM counterparts, especially in the accretion disk near a supermassive black hole [50, 51]. Detecting these EM counterparts is crucial as it extends GW astronomy to multi-messenger astronomy, enabling us to gather a wealth of physical information from different observations.

Throughout this thesis, I will focus on compact binaries, but I will also provide a brief overview of other astrophysical sources in the following sections.

1.2.2 Pulsars

Rotating systems are not limited to binaries. Pulsars are dense, highly magnetized, and fast-rotating neutron stars that emit EM radiation out of their magnetic poles. Pulsars are not perfect spheres, and they have a non-zero changing mass quadrupole due to their irregularity and rotation, which produces GWs [52].

There are different types of irregularity of pulsars and they generate distinct GW signals. The first type is triaxial irregularity: a neutron star may have "bumps" or "mountains" from formation, magnetic field, stresses due to relaxation and many other reasons [53, 54, 55]. This persistent irregularity, combined with the star's rotation, produces long-lasting *continuous* GWs. Unlike the CBC chirp signals, continuous waves from isolated pulsars decay over time because pulsars' rotation gradually slows down ("spin-down") due to energy loss by gravitational radiation, and the frequency evolution is much slower than compact binaries. Typical continuous GW magnitude from isolated rotating pulsars is [6]

$$h \sim 1.06 \times 10^{-25} \left(\frac{\epsilon}{10^{-6}}\right) \left(\frac{I_{zz}}{10^{38}\ \mathrm{kg\ m^2}}\right) \left(\frac{10\mathrm{kpc}}{r}\right) \left(\frac{f}{1\mathrm{kHz}}\right)^2, \tag{1.100}$$

where ε is the equatorial ellipticity of the pulsar and I_{zz} is principle moment of inertia. Continuous GW is weaker than GW from compact binaries (Eq. 1.96). However, the signal accumulates with observation time faster than the noise, therefore it is possible to detect continuous waves at some point. Indeed, even a null result is valuable as it imposes constraints on the existing electromagnetically observed neutron stars [56]. In some cases, the rotation induces precession, which potentially produces long or medium-term GW emissions [6, 57].

The second type of irregularity is the vibrations of pulsars. Pulsars may go through various vibrational modes, including f-modes (fundamental modes), the first resonant mass quadrupolar mode, excited by glitches or nuclear explosion or other reasons, and R-modes (Rossby waves), the mass current quadrupole in hot young or accreting neutron stars, and many other modes [52, 58]. Some of these vibrations are short-lived (like f-modes) and produces short *burst* GW emissions. Some are long-lived, like R-modes, produces continuous waves.

All the discussions about pulsars above are also applicable to other relativistic stars like boson stars, if they do exist. The observation of GWs is an ideal test of such exotic objects, as they are predicted to exhibit different equatorial ellipticity ε compared to neutron stars, which can be estimated by GW observations.

1.2.3 Gravitational collapse

Neutron stars and black holes are produced by the gravitational collapse of massive and old stars or the core collapse of accreting white dwarfs. If the collapse is nonspherical and driven by strong rotation, it may generate burst GWs that carry away binding energy and angular momentum [59, 60]. Despite extensive efforts to simulate the collapse process [59, 61, 62, 63], fully understanding such events remains challenging due to complexities in hydrodynamics, neutrino physics, realistic nuclear physics, magnetic fields, and other factors. It is therefore impossible to have a closed-form waveform equation like in CBC (chirp signals) and continuous waves (quasi-monochromatic signals). Therefore, detecting burst GWs from gravitational collapse requires unmodelled searches that only make use of the coherence between different detectors.

Simulations reveal that 99% of energy in this process is carried away by neutrinos, and most of the rest of the energy is converted to the kinetic energy of the explosion, while the EM (this process is also accompanied by GRB [64]) and GW only get 0.01% [65]. GWs from a typical supernova collapse might extract between approximately 10^{-7} and 10^{-5} of the total available mass-energy, and the frequency bands might lie in 20-200Hz [66, 67]. The typical burst GW amplitude is roughly given by [60]

$$h \sim 6 \times 10^{-21} \left(\frac{E}{10^{-7} M_\odot}\right)^{1/2} \left(\frac{1\text{ ms}}{T}\right)^{1/2} \left(\frac{1\text{kHz}}{f}\right) \left(\frac{10\text{kpc}}{r}\right), \tag{1.101}$$

where E is the equivalent mass-energy in $M_\odot$ and T is the duration of the burst. This amplitude is large at 10kpc, but the event rate within 10kpc brings more uncertainties to an early detection.

1.2.4 Stochastic background

We have discussed discrete sources in the past sections. Other than that, the universe has a random GW background resulting from the superposition of all discrete systems that emit GWs and from some cosmological process, in analogy to the Cosmic Microwave Background (CMB). The inspiralling binaries we mentioned in Sec. 1.2.1 are an example of an astrophysical stochastic background. Various phenomena in the early universe, such as inflation [68], preheating [69], phase transitions [70], and cosmic strings [71] are predicted to contribute to a primordial stochastic background. The stochastic background is a probe of both galactic binaries and the early universe.

The stochastic background needs to be quantified statistically. It spans over all frequency bands, ranging from nanohertz to kilohertz, so it is conventional to talk about the energy density per unit logarithm of frequency, and after averaging over all source

directions and all independent polarizations, the energy density is defined as [60]

$$\frac{d\rho_{\rm gw}}{d\ln f} = 4\pi^2 f^3 S_{\rm GW}(f), \tag{1.102}$$

where $S_{\rm GW}(f)$ is the power spectrum of the stochastic background. Its ratio over the critical density $\rho_{\rm c}$ is defined as $\Omega_{\rm GW}$, in analogy to $\Omega_{\rm m}$ and Ω_Λ in cosmological models:

$$\Omega_{\rm GW}(f) = \frac{1}{\rho_{\rm c}}\frac{{\rm d}\rho_{\rm GW}}{{\rm d}\ln f} = \frac{4\pi^2}{3H_0^2} f^3 S_{\rm GW}(f), \tag{1.103}$$

where H_0 is the Hubble constant. The $\Omega_{\rm GW}$ is expected to be very small ($< 10^{-5}$), as it is otherwise in tension with the standard Big Bang model of nucleosynthesis [60].

The stochastic background is indistinguishable from noise in a short period in one detector. In the long term, if the detector changes its orientation, it is in principle possible to distinguish the background from the noise [6]. A more promising way of detecting the stochastic background is using the cross-correlation between different detectors [72], as they have different responses to the background GWs.

1.3 Gravitational wave detectors

The first attempt at GW detection can be traced back to the 1960s when Joseph Weber built a *resonant bar* at the University of Maryland [73]. The resonant bar gets excited and produces vibrations in response to GWs due to the changes in the stress tensor inside the bar. Weber claimed the first detection of signals from the center of the Galaxy in 1969 [74], but subsequent attempts by other research groups, achieving similar or better sensitivity, failed to reproduce his results [75, 76, 77, 78, 79]. Weber's detection was therefore not recognized by the science community, and the research on resonant bar detectors gradually waned.

An indirect detection of GWs was later made by Hulse and Taylor in 1974 [8]. They observed an inspiralling BNS system and recorded the decay of its orbit period. The changes in orbit period matched the prediction by GR perfectly and thus verified the energy loss of this system was consistent with the energy carried away by gravitational radiation. For this groundbreaking discovery of the binary pulsar and the indirect evidence of gravitational waves, they were awarded the Nobel Prize in Physics in 1993.

Using *interferometers* as GW detectors was first proposed in the 1960s [80]. After decades of collaborative efforts by an international team and several observing runs [81, 82, 83], a milestone was achieved in 2015 when the Advanced Laser Interferometer Gravitational-Wave Observatory (aLIGO)[4] directly detected GWs from the merger of a binary black hole

[4] aLIGO is the upgraded version of the original LIGO. I will not distinguish them in the subsequent

system [7] with dual 4km-long interferometers. This historic event opened the window to GW astronomy. Three leaders of LIGO, Rainer Weiss, Barry Barish, Kip Thorne have been awarded the Nobel Physics Prize in 2017.

While ground-based detectors continue to detect new CBCs, Pulsar Timing Array (PTA) community claimed the detection of stochastic background in 2023 at the nanohertz band [84, 85, 86, 87], marking another breakthrough in GW astronomy.

In this section, I will introduce these GW detectors, focusing primarily on the ground-based interferometers, including their response, noise, current status and future prospects. I will also provide a brief overview of new types of gravitational wave detectors at the end of this section.

1.3.1 Ground-based interferometers

Configurations of interferometers

Interferometer GW detectors are based on Michelson interferometers, as shown in Fig. 1.4. The laser is diverted into two equal-length arms by a beam splitter and travels back to the photon detector below. As discussed in Sec. 1.1.2, passing GWs alter the proper distance along the arms, so the optical distance of the laser changes, leaving different interfering patterns to the photodetector, which can be later interpreted as GW readout. The actual LIGO setup, however, is more complex than this basic setup. Several critical components are employed to enhance the detector's sensitivity:

- Seismic Isolation. The hanging system of the mirror (test mass) consists of a complex quadruple-pendulum system that could passively and actively remove the seismic vibrations and make sure the mirror is static [88, 89, 90].

- Strong and Stable Laser. LIGO's laser system generates high-power near-infrared laser to improve the precision of measurement. A series of feedback mechanisms are used to stabilize the power fluctuations of the laser before the laser is used in the interferometer [91].

- Ultra-High Vacuum. LIGO's optical components are in an ultra-high vacuum environment, approximately one trillionth of an atmosphere [92], minimizing laser scattering with air molecules

- Ultra-Reflective and Stable Mirrors. The mirrors are made highly reflective to prevent photon absorption and heating - it only absorbs one out of every 3.3 million photons [93]. The mirror contains dozens of layers of optical coatings to achieve nanometer smoothness [94]. An auxiliary system using a CO_2 laser is used to precisely counteract the shape changes due to heating [89].

text.

- Modecleaner Cavities. Modecleaner cavities are added after the laser output and before the photodetector to filter input and output fields [88]. This helps reduce the laser noise and improve the beam quality.
- Arm Cavities. Each arm includes a Fabry-Perot cavity which allows the laser to move back and forth thousands of times [88]. This could increase the circulating power and lengthen the effective arm length.
- Power Recycling. A power recycling system (a mirror is added to form a resonant cavity) is put before the beam splitter to increase the input laser power by a factor of $\mathcal{O}(1000)$ [95, 96].
- Signal Recycling. A mirror is added before the photodetector to recycle the laser back to the interferometer [97, 98]. It allows adjustment of the maximum sensitivity band and improves the SNR.
- Light Sqeezer. Laser is injected in squeezed states to beat the standard quantum limit [99, 100, 101], which achieves an improved sensitivity at some frequency bands, especially at frequencies less than 150Hz.

As LIGO detectors are extremely complex and under fast development, this overview cannot comprehensively cover all modules and techniques utilized. More details can be found in, e.g., Refs. [88, 92, 98]. In addition to LIGO, credit must also be given to the Virgo collaboration [102] and KAGRA collaboration [103]. They joined GW search since O2 in 2017 and O4 in 2023, respectively.

Response to GWs

From this point onward, we will consider the simplified Michelson structure illustrated in Fig. 1.4 and discuss its response to GWs following Whelan [104]. Assume the unit vectors along the detector arms are u_i, v_i, and two arms have the same length L_0. When GW passes through, and $h_{\mu\nu}$ is given in TT gauge:

$$h_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}_{\mu\nu}, \tag{1.104}$$

we will have the following ds^2

$$ds^2 = -c^2dt^2 + (1+h_+)dx^2 + (1-h_+)dy^2 + 2h_\times dxdy + dz^2, \tag{1.105}$$

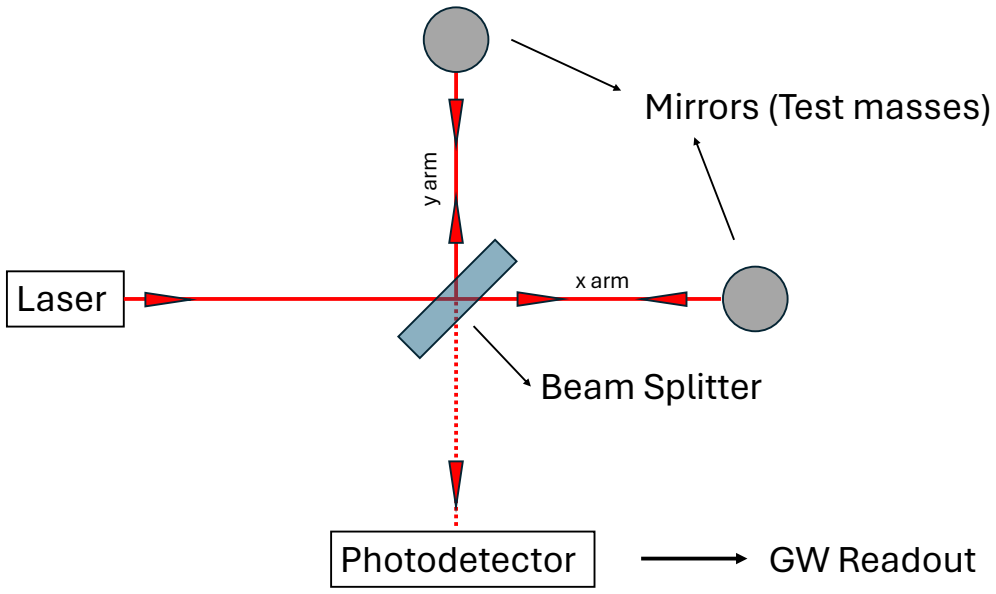


Figure 1.4: The structure of a basic Michelson interferometer.

We assume $h_+, h_\times$ are constant over the lengthscale L_0, i.e., the wavelength of GW is much longer than the detector armlength. This is a good approximation for LIGO-like ground-based detectors whose arm is about kilometers long. $\lambda_{\rm GW} \gg 4km$ means $f_{\rm GW} \ll 75$kHz, this is a safe limit given the frequency bands of the astrophysical sources in Sec. 1.2. We previously computed the proper distance for the plus polarization alone in Eq. 1.29, and similar calculations can be applied to determine the proper distance here. Consider the proper distance along arm $u_i = (u_1, u_2, u_3)$:

$$\begin{aligned} ds_u^2 &= L_0^2 \left[(1+h_+)u_1^2 + (1-h_+)u_2^2 + u_3^2 + 2h_\times u_1 u_2\right] \\ &= L_0^2 \left(1 + h_{11}u_1^2 + h_{22}u_2^2 + h_{12}h_{21}u_1u_2\right) \\ &= L_0^2 \left(1 + u^i u^j h_{ij}\right), \end{aligned} \tag{1.106}$$

and therefore

$$s_u = L_0 \left(1 + u^i u^j h_{ij}\right)^{1/2} \approx L_0(1 + \frac{1}{2}u^i u^j h_{ij}). \tag{1.107}$$

The interferometer measures the difference in roundtrip times between the two arms, so the measurable quantity is the *strain* h defined by[5]

$$h = \frac{s_u - s_v}{L_0} = h_{ij}\frac{u^i u^j - v^i v^j}{2} = D^{ij}h_{ij}, \tag{1.108}$$

[5]In fact, what the photodetector record is the voltage $v(f)$, and the strain data is reconstructed by the voltage and a response function $R(f)$ that relates the digital readout and GW strain, i.e., $d(f) = R(f)v(f)$ [105]. Here d is the detector data, which equals signal h plus noise n.

where

$$D^{ij} = \frac{u^i u^j - v^i v^j}{2} \tag{1.109}$$

is called *detector tensor*. That is to say, the response of an interferometer to GW is the *projection* of the GW polarizations onto the detector tensor. To further simply this equation, write $h_{ij} = h_+(C_+)_{ij} + h_\times(C_\times)_{ij}$, where C_+ and $C_\times$ are polarization tensors in Eq. 1.27. We have

$$h = D^{ij}(h_+(C_+)_{ij} + h_\times(C_\times)_{ij}) = F_+ h_+ + F_\times h_\times, \tag{1.110}$$

where

$$F_+ = C_+^{ij} D_{ij},\ F_\times = C_\times^{ij} D_{ij} \tag{1.111}$$

are *antenna response functions* of the interferometer. The above equation is written in the detector frame, i.e., the x and y axes are along the detector arms. In practice, the waveform polarizations are calculated in the source frame, so a series of coordinate transformations are required. There are two equivalent formalisms of calculating the response: the first one transforms all tensors into the detector coordinate [106], while the second transforms all tensors into the fixed Earth coordinate [107]. We will adopt the second one here as it is widely used in GW-relevant software (e.g. `LALSuite` [108], `Bilby` [109]) and is convenient to perform parallel calculations (see, e.g., Sec. 6.4.1).

The geometry is set up in Fig. 1.5. The source frame (x, y, z) (with only the z axis shown) and the wave frame (x', y', z') correspond to the same coordinates with the same names as in Fig. 1.2. The z' axis points towards the Earth. The z and z' axes are naturally fixed[6], while the freedoms of choosing (x, y) and (x', y') are usually fixed for convenience in physics (waveform calculation), and an initial phase ϕ_0 will be given in this process[7]. The celestial coordinate is (x'', y'', z''), and right ascension α and declination δ are defined in this frame. x'' points to the vernal equinox. The sky direction of the source in the celestial coordinate is (α, δ). The wave frame x' axis has an angle with the direction of decreasing right ascension, and this angle is called *polarization angle* ψ. This angle defines GW polarization tensors in the wave frame. When $\iota = 0$, it is degenerate with ϕ_0 (or equivalently, with the coalescence phase ϕ_c).

Returning to Earth, there is a fixed Earth coordinate (x''', y''', z''') that shares the z'' axis with the celestial frame. This is the frame in which we want to calculate response functions. Let x''' point to the prime meridian and let y''' point to the $90°E$. We use longitude φ and latitude λ to describe the location of the vertex of the detector.

The detector frame is denoted as (x'''', y'''', z''''). We use azimuth angle ψ_x and ψ_y

[6] We do not consider source precession here. If precession occurs, z can be chosen to be the direction of total angular momentum instead of the orbital angular momentum.

[7] For eccentric orbit, the orbit eccentricity and eccentric anomaly will also be defined [110].

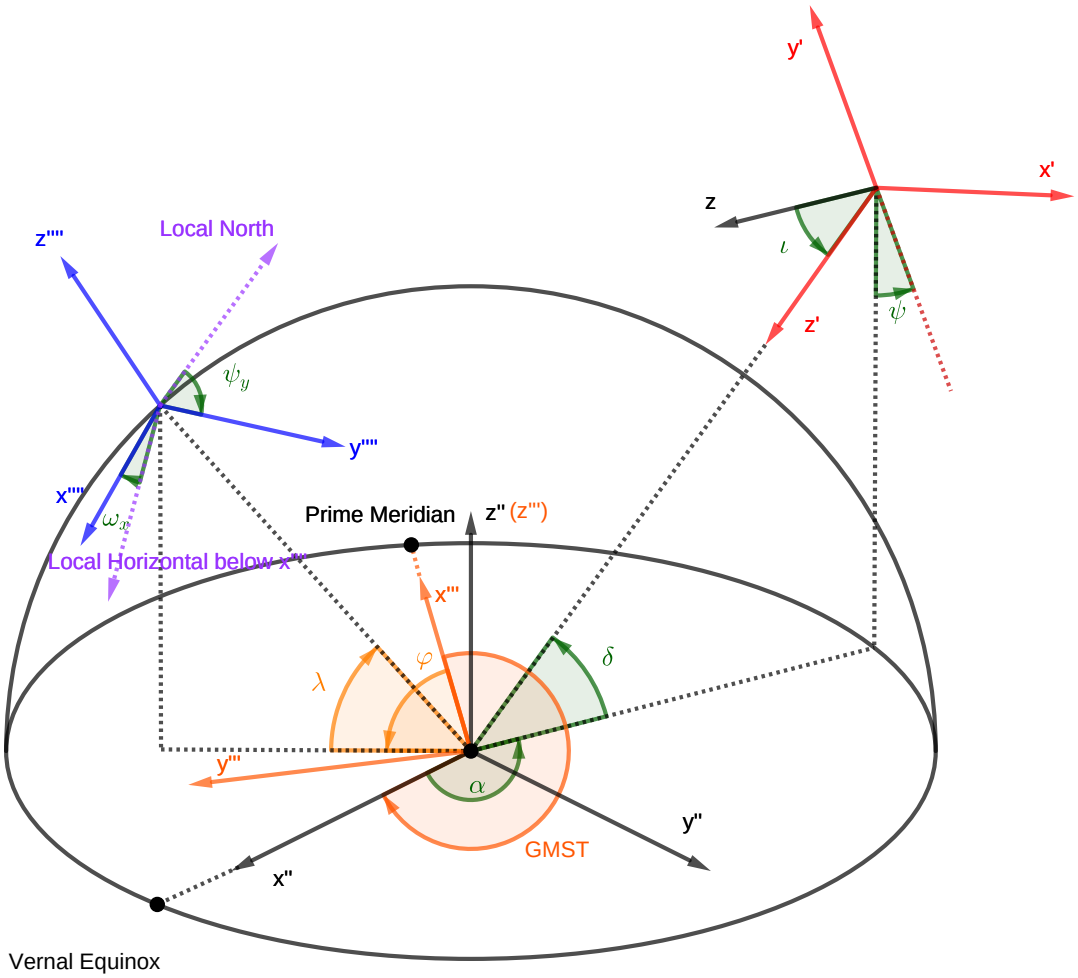


Figure 1.5: Different coordinates for calculating the detector response. (x, y, z): source frame (only z axis shown). (x', y', z'): wave frame. (x'', y'', z''): celestial coordinate, whose x'' points to the vernal equinox. (x''', y''', z'''): fixed Earth coordinate, whose x''' points to the prime meridian. (x'''', y'''', z''''): detector frame. The detector arms are along x'''' and y''''. More details in the text.

(measured from the local north to the east) to describe the orientation of the detector, and we use the tilt angle ω_x and ω_y to describe the arms' tilt from the horizontal. The tilt angle is usually very small ($\sim 10^{-5} - 10^{-4}$rad) but we exaggerated it in Fig. 1.5. We did not label ψ_x and ω_y for the clarity of the plot.

We first calculate the polarization tensors in the fixed Earth coordinate (x''', y''', z'''). Consider a standard spherical coordinates (θ, ϕ) defined in (x''', y''', z'''), based on Fig. 1.5, we have $\theta = \pi/2 - \delta$ and $\phi = \alpha - \text{GMST}$. Greenwich Mean Sidereal Time (GMST) is measured from x''' (prime meridian) to the vernal equinox by definition. The matrix

transforming (x', y', z') to (x''', y''', z''') is given by [106]

$$M = \begin{pmatrix} \sin\phi\cos\psi - \cos\phi\sin\delta\sin\psi & -\cos\phi\cos\psi - \sin\phi\sin\delta\sin\psi & \cos\delta\sin\psi \\ -\sin\phi\sin\psi - \cos\phi\sin\delta\cos\psi & \cos\phi\sin\psi - \sin\phi\sin\delta\cos\psi & \cos\delta\cos\psi \\ -\cos\phi\cos\delta & -\sin\phi\cos\delta & -\sin\delta. \end{pmatrix} \tag{1.112}$$

The basis vectors in the wave frame $\hat{x}'_i, \hat{y}'_i$ are transformed to the fixed Earth frame by $\hat{X}'''_i = M_{ij}\hat{x}'_j$, $\hat{Y}'''_i = M_{ij}\hat{y}'_j$. Given that $\hat{x}'_i = (1,0,0)_i$ in the wave frame, the transformed basis vector simply takes the first column from M_{ij} (and similar for $\hat{y}'_i$):

$$X'''_i = M_{i1},\ Y'''_i = M_{i2}. \tag{1.113}$$

The polarization tensor is then defined as

$$C^{+'''}_{ij} = X'''_i X'''_j - Y'''_i Y'''_j,\ C^{\times'''}_{ij} = X'''_i Y'''_j + Y'''_i X'''_j. \tag{1.114}$$

Now we compute the detector tensor. The Cartesian coordinates of the detector in the fixed Earth frame is [107]

$$r'''_i = \left([R(\varphi)+h]\cos\varphi\cos\lambda,\ [R(\varphi)+h]\cos\varphi\sin\lambda,\ [(b^2/a^2)R(\varphi)+h]\sin\varphi\right)_i, \tag{1.115}$$

where $R(\varphi) = a^2\left(a^2\cos\varphi + b^2\sin\varphi\right)^{-1/2}$ is the local radius of the Earth, with $a = 6378137$m and $b = 6356752.314$m based on the WGS-84 Earth Model. h is the elevation of the detector site. At this position, the unit vectors pointing East, North, and Up are

$$\begin{aligned} \hat{\lambda}_i &= (-\sin\lambda,\ \cos\lambda,\ 0)_i \\ \hat{\varphi}_i &= (-\sin\varphi\cos\lambda,\ -\sin\varphi\sin\lambda,\ \cos\varphi)_i \\ \hat{h}_i &= (\cos\varphi\cos\lambda,\ \cos\varphi\sin\lambda,\ \sin\varphi)_i, \end{aligned} \tag{1.116}$$

in the fixed Earth frame, and the x-arm of the detector is given by

$$u'''_i = \cos\omega_x\cos\psi_x\hat{\lambda}_i + \cos\omega_x\sin\psi_x\hat{\varphi}_i + \sin\omega_x\hat{h}_i, \tag{1.117}$$

and v'''_i have the same equation with the subscript change $x \to y$. The detector tensor D'''_{ij} can then be calculated:

$$D'''_{ij} = \frac{u'''_i u'''_j - v'''_i v'''_j}{2}, \tag{1.118}$$

combining Eq. 1.114, the response functions can be obtained by

$$F_{+,\times}(\alpha,\delta,\psi,\text{time}) = D'''_{ij}C^{'''ij}_{+,\times}(\alpha,\delta,\psi,\text{time}). \tag{1.119}$$

In addition to amplitude projection, a time shift must be applied if a universal time

parameter is used to describe the GW event (rather than local times for each detector). For example, using the time that GW arrives at the Earth center t_c, the time shift is given by

$$\Delta t = t_{\rm det} - t_c = r'''_i(n''')^i/c, \tag{1.120}$$

where c is the speed of light and

$$n'''_i = (-\cos\phi\cos\delta,\ -\cos\delta\sin\phi,\ -\sin\delta)_i \tag{1.121}$$

is the wave vector in the fixed Earth frame describing the direction of GW propagation, pointing from the source to the Earth. Given the Earth's radius of 6371km, Δt should always be less than 22ms. The response of a detector is therefore

$$h_{\rm det}(t) = h(t+\Delta t) = F_+(\alpha,\delta,\psi,t)h_+(t+\Delta t) + F_\times(\alpha,\delta,\psi,t)h_\times(t+\Delta t), \tag{1.122}$$

and the Fourier transform goes as [111]

$$\mathcal{F}\left[h(t+\Delta t)\right] = e^{-2\pi i f\Delta t}\mathcal{F}\left[h(t)\right]. \tag{1.123}$$

If the time duration of the signal is short, response functions can be considered unchanged as the Earth barely rotates during the GW event, so t in the response functions can be replaced with the geocenter time t_c. The time delay from the geocenter (or equivalently, the time delay between different detectors) is crucial to the sky localization of GW sources, as the sky triangulation information is encoded in the time delay [112].

The detailed parameters (locations, orientations, etc) for LIGO, Virgo, and KAGRA and many other previous or proposed detectors can be found in `LALDetectors.h`[8] in `LALSuite` [108].

1.3.2 Characterizing noises

Noise sources

GW experiment is one of the most precise experiments in the world, and understanding and characterizing noises is crucial to successful detection. The GW data d consists of signal h (which is calculated in the previous section) and noise n, i.e.,

$$d = h + n, \tag{1.124}$$

and this equation holds in both time and frequency domains. The possible noise sources are [88]

[8] https://lscsoft.docs.ligo.org/lalsuite/lal/_l_a_l_detectors_8h_source.html

- Seismic noise. Ground-based GW detectors are sensitive to all kinds of seismic vibrations, including solid Earth tides (< 10Hz), microseismic peaks (~ 160mHz) generated by water waves pushing on the ocean floor, and trains and cars nearby. Isolation systems and active noise cancellation systems are used to manage the seismic noise.

- Quantum noise. Noise exists even in the absence of laser amplitude due to the quantum nature of the laser. This includes shot noise, the noise from error in photon counting, which corresponds to the uncertainty in phase measurement, and radiation pressure noise, the back-action noise caused by the random motion of the mirrors driven by the fluctuations in laser power, which corresponds to the uncertainty in amplitude. The former dominates high frequencies and the latter dominates the low. Squeezed light is used to correlate these two uncertainties and achieve a surpass-quantum-limit sensitivity in some frequency band [100].

- Gravitational gradient noise (Newtonian noise). The gravitational interaction between moving masses and the free test mass (mirror). It can be caused by density perturbations due to vehicles, clouds or seismic surface waves (S-waves). It is hard to shield the test mass from gravitational gradient noise as we can not screen gravity. Future detectors may go underground to reduce the effect of this noise source to improve the low-frequency sensitivity.

- Coating Brownian noise (Test mass thermal noise I). Arising from the mechanical dissipation in the coatings. Parts of the system close to the laser beam contribute more to thermal noise than others, so the optical coatings play a dominating role. This could happen at the most sensitive frequencies (~ 100Hz) of ground-based GW detectors.

- Coating thermal-optic noise (Test mass thermal noise II). Arising from the fluctuations in the temperature of the test mass. This includes thermo-elastic noise: temperature fluctuations are coupled to the test mass displacement measurement via the coefficient of thermal expansion, i.e., local regions of the mirror can expand or contract, changing the optical distance. This also includes thermo-refractive noise: temperature fluctuations change the refractive index of coatings, bringing a phase change in the laser, which looks like a displacement of the test mass.

- Substrate Brownian noise (Test mass thermal noise III). Thermal energy drives resonant modes of suspended mirrors, and the width of the resonance is related to the mechanical loss of the test mass substrate material (fused silica). Here mechanical loss refers to energy dissipation by internal friction in material.

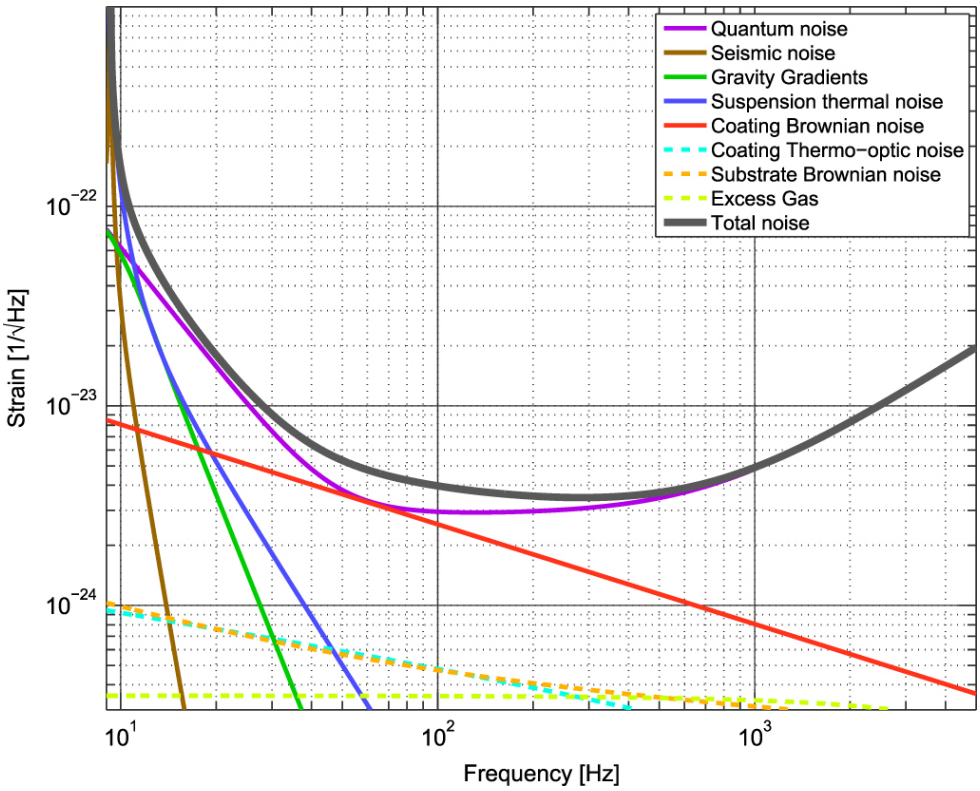


Figure 1.6: Principal noise terms for aLIGO. Credit: Ref. [88].

- Suspension thermal noise. The thermally excited motion of the suspension fibers that hang the mirror. This includes various vibration modes: pendulum mode, violin mode, and Vertical bounce mode.

- Residual gas noise. The detector arm tube is not a perfect vacuum. Residual gas in the tubes will lead to statistical variations in the column density of gas particles in the beam path, producing fluctuations in the effective refractive index along the path.

- Other noises during operation, such as ravens in the desert of Hanford [113].

The noise is quantified by the *power spectrum density* $S_n(f)$, or PSD, defined as [114]

$$\langle \tilde{n}(f)\tilde{n}\,(f')\rangle = \frac{1}{2}S_n(f)\delta\,(f-f')\,, \tag{1.125}$$

where $\langle \dots \rangle$ denotes ensemble average. $\tilde{n}(f)$ is the Fourier transform of time domain noise $n(t)$. For negative frequencies, it should be calculated as $S_n(|f|)$. The square root of PSD is Amplitude spectral Density (ASD). The GW signal should at least reach a similar order of magnitude of ASD to be detected. An illustration of the contributions of different noise sources to the noise amplitude spectrum of aLIGO is given in Fig. 1.6. The most sensitive band for LIGO is 20-2000Hz.

Statistical model of the detector noise

The statistical noise model can be built upon the assumption that the noise is *stationary* and *Gaussian* [111]. Gaussian means that the noise $n(t)$ is sampled from a Gaussian process with zero mean, and stationary requires the noise characteristics not to change with time (at least for a short period). In reality, neither of these assumptions is strictly true. There are glitches in the detector that break down the Gaussianity, and the noise level fluctuates during operation [115]. However, the stationarity is a reasonable approximation for current detectors as the signals are not very long due to the limit by the low-frequency noise. Glitches are one of the primary challenges to data quality: they appear frequently in the data and sometimes overlaps with true signals [115, 116, 117]. Significant efforts have been made to identify and remove glitches from the data, e.g. [118, 119, 120, 121, 122, 123].

Now we construct the statistical noise model. Consider a series of noise samples $\{n_i\}$ sampled at $\{t_i\}$ within duration T, $i = 1, 2, \ldots, N$. We have $\Delta t = T/(N-1)$ and $t_j - t_k = (j-k)\Delta t$. Since we assume zero mean, the correlation function and covariance function are the same:

$$C(t_i, t_j) = \langle n(t_i) n(t_j) \rangle \,, \tag{1.126}$$

and the covariance matrix can be defined as:

$$C_{ij} = C(t_i, t_j). \tag{1.127}$$

The distribution of the noise $n = \{n_i\},\ i = 1, 2, \ldots, N$ is the multivariate Gaussian:

$$P(n) = \frac{e^{-\frac{1}{2} C^{-1}_{ij} n_i n_j}}{[(2\pi)^N |\det C_{ij}|]^{1/2}}, \tag{1.128}$$

where C^{-1}_{ij} is the inverse of the covariance matrix, i.e., $C_{ij} C^{-1}_{jk} = \delta_{ik}$. Here repeating index represents summation and we do not distinguish between upper and lower indices. Eq. 1.128 can be interpreted as "the probability of seeing the noise realization $\{n_i\},\ i = 1, 2, \ldots, N$". In principle, the statistical model is already built by Eq. 1.128, but the correlation function in the time domain is expensive to compute in practice. We need to find a simpler representation.

Given the noise is stationary, the correlation between t_i and t_j is only relevant to the difference between t_i and t_j. This means that, we can use the autocorrelation function $C_n(\tau)$ to describe the stochastic process. For $\tau = t_i - t_j$, we have

$$C_n(\tau) = C(t_i, t_j). \tag{1.129}$$

The Wiener-Khintchine theorem (e.g., Section 28 in [124]) tells us the PSD is the Fourier

transform of the autocorrelation function:

$$S_n(f) = 2\int_{-\infty}^{\infty} C_n(\tau)e^{-2\pi i f\tau}d\tau. \tag{1.130}$$

The denominator in Eq. 1.128 is a normalization constant that can be factored out, so we will focus on the numerator. In the continuum limit ($\Delta t \to 0,\ T \to +\infty$), the exponent in Eq. 1.128 can be rewritten by changing summation to integral:

$$\begin{aligned}
C^{-1}_{jk}n_j n_k &= \frac{1}{\Delta t^2}C^{-1}_{jk}n_j n_k \Delta t\Delta t \\
&= \frac{1}{\Delta t^2}\iint C^{-1}(t_j,t_k)n(t_j)n(t_k)dt_j dt_k \\
&= \frac{1}{\Delta t^2}\int n(t_k)dt_k \int C^{-1}(t_j,t_k)n(t_j)dt_j \\
&= \frac{1}{\Delta t^2}\int n(t_k)dt_k \int \left[\int C^{-1}(t_j,t_k)e^{-2\pi i f t_j}dt_j\right]\tilde{n}^*(f)df,
\end{aligned} \tag{1.131}$$

where $C^{-1}(t_j,t_k) = C^{-1}_{jk}$ and superscript $*$ denotes complex conjugate. The integral limits are all infinity and are omitted in the equations. From the third to the last line, we have used the Parseval-Plancherel identity: given two functions $a(t)$ and $b(t)$ and their Fourier transform $\tilde{a}(f)$ and $\tilde{b}(f)$, we have

$$\int_{-\infty}^{+\infty} a(t)b^*(t)dt = \int_{-\infty}^{+\infty} \tilde{a}(f)\tilde{b}^*(f)df. \tag{1.132}$$

We then calculate the bracket in the last line of Eq. 1.131:

$$\begin{aligned}
\int C^{-1}(t_j,t_k)e^{-2\pi i f t_j}dt_j &= \frac{2\int C_n(\tau)e^{-2\pi i f\tau}d\tau}{S_n(f)}\int C^{-1}(t_j,t_k)e^{-2\pi i f t_j}dt_j \\
&= \frac{2}{S_n(f)}\int C(t_l,t_j)e^{-2\pi i f(t_l-t_j)}dt_l \int C^{-1}(t_j,t_k)e^{-2\pi i f t_j}dt_j \\
&= \frac{2}{S_n(f)}\iint C(t_l,t_j)C^{-1}(t_j,t_k)e^{-2\pi i f t_l}dt_l dt_j \\
&= \frac{2\Delta t^2}{S_n(f)}C_{lj}C^{-1}_{jk}e^{-2\pi i f t_l} \\
&= \frac{2\Delta t^2}{S_n(f)}\delta_{lk}e^{-2\pi i f t_l} \\
&= \frac{2\Delta t^2}{S_n(f)}e^{-2\pi i f t_k}.
\end{aligned} \tag{1.133}$$

From the first line to the second line, we used the fact that t_j can be any value since t_l

runs from $-\infty$ to $+\infty$. Substituting this into Eq. 1.131, we get

$$\begin{aligned} C_{jk}^{-1} n_j n_k &= \int n(t_k) dt_k \int \frac{2}{S_n(f)} e^{-2\pi i f t_k} \tilde{n}^*(f) df \\ &= \int \frac{2}{S_n(f)} \tilde{n}^*(f) df \int n(t_k) e^{-2\pi i f t_k} dt_k \\ &= 2 \int_{-\infty}^{+\infty} \frac{\tilde{n}^*(f)\tilde{n}(f)}{S_n(f)} df \\ &= 4 \int_{0}^{+\infty} \frac{\tilde{n}^*(f)\tilde{n}(f)}{S_n(f)} df \\ &= 4 \sum_k \frac{\tilde{n}_k \tilde{n}_k^*}{T S_k} \\ &= \sum_k \frac{\mathrm{Re}(\tilde{n}_k)^2 + \mathrm{Im}(\tilde{n}_k)^2}{T S_k/4}. \end{aligned} \tag{1.134}$$

We adjusted the integral limit to $f > 0$ to be consistent with the definition of $S_n(f)$[9]. To discretize the equation we used $\Delta f = 1/T$, $\tilde{n}_k = \tilde{n}(f_k)$, $S_k = S_n(f_k)$. Re and Im denote the real and imaginary parts, respectively. Eq. 1.134 implies that, in the frequency domain, the noise's real and imaginary components also follow the Gaussian distribution with zero mean and variance of $TS_k/4$:

$$\tilde{n}_k \sim N\left(0, \frac{TS_k}{4}\right) + iN\left(0, \frac{TS_k}{4}\right). \tag{1.135}$$

Most importantly, unlike the correlated time domain noise (i.e., non-diagonal covariance matrix C_{ij}), the frequency bins are independent and the covariance matrix in the frequency domain is $\mathrm{diag}(TS_k/4)$. This property significantly accelerates the likelihood evaluation in the frequency domain compared to the time domain.

The normalization factor can be written as the product of normalization factors of real and imaginary parts, which results in $\prod_k \frac{\pi T S_k}{2}$. Putting the equations together, we can finally write down the likelihood from Eq.1.128 to

$$P(n) = \frac{1}{\prod_k \frac{\pi T S_k}{2}} e^{-\frac{1}{2}\sum_k \frac{4\tilde{n}_k \tilde{n}_k^*}{T S_k}} \propto e^{-\frac{1}{2}(n|n)}, \tag{1.136}$$

where the $(n|n)$ is the *noise-weighted inner product* defined as:

$$(a|b) = 4\mathrm{Re} \int_0^{+\infty} \frac{\tilde{a}^*(f)\tilde{b}(f)}{S_n(f)} df = 4\mathrm{Re} \sum_k \frac{\tilde{a}_k^* \tilde{b}_k}{T S_k}. \tag{1.137}$$

Note there are other equivalent definitions of inner product in the literature, and I may

[9]Since noise is real-valued in the time domain, it is symmetric in the frequency domain: $\tilde{n}(f) = \tilde{n}^*(-f)$. On the other hand, $S_n(f) = S_n(-f)$, so the integral can be expressed in $f > 0$ regime.

use them without specifying:

$$\begin{aligned}(a|b) &= 4\mathrm{Re}\int_0^{+\infty}\frac{\tilde{a}^*(f)\tilde{b}(f)}{S_n(f)}df \\ &= 4\mathrm{Re}\int_0^{+\infty}\frac{\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)}df \\ &= 2\int_0^{+\infty}\frac{\tilde{a}^*(f)\tilde{b}(f)+\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)}df \\ &= \int_{-\infty}^{+\infty}\frac{\tilde{a}^*(f)\tilde{b}(f)+\tilde{a}(f)\tilde{b}^*(f)}{S_n(f)}df\end{aligned} \tag{1.138}$$

1.3.3 Current detection status and future prospects

Since the first GW detection GW150914 [7], the LIGO-Virgo-KAGRA collaboration (LVK) has been upgrading the detectors and has performed three observing runs (O1 [125],O2 [125], and O3 [126, 127, 128]). The O4 is ongoing as of the time of writing this thesis. Two aLIGO detectors, H1 and L1 at Hanford and Livingston, respectively, have participated in all observing runs since the first detection. Their detection range[10] are improved from approximately 60 Mpc at the beginning of O1 [125] to approximately 160 Mpc at the beginning of O4 [129]. Virgo detector (V1) [102] joined O2 in 2017 with a horizon of approximately 25 Mpc [125], which was improved to 56 Mpc by the end of O3. KAGRA [128] (the successor of TAMA300 [130]) and GEO-600 [131] joined O3 in 2020 and achieved horizon around 1 Mpc [132].

To date, over 100 CBC events have been detected by the LVK, including several remarkable events. GW170817 [116] is the first BNS event and it is associated with a strong GRB [133] and neutrinos [134]. The sky localization of GW170817 was provided hours after the merger, based on which scientists all over the world carried out a successful multi-band EM follow-up [49], marking the beginning of multi-messenger astronomy. Properties of neutron stars and kilonova are thoroughly investigated with the help of GW170817 and its EM follow-up [135, 136]. GW190521 [137] is the first BBH event whose remnant black hole mass is above 100 $M_\odot$, revealing hints of intermediate-mass black hole formation. GW190814 [138] and GW230529 [129] have a component that lies in the mass gap between neutron stars and black holes, indicating the existence of exotic stars or a different population of known objects. Many intriguing phenomena are reported, including higher modes of GWs [138, 139], precession [140], orbit eccentricity [141, 142], overtones [143, 144, 145, 146], recoil of the remnant [147] and so on. On the population level, these CBC signals are used to reconstruct compact binary popula-

[10]Range of a detector is defined as the sky and orientation averaged detection luminosity distance of $1.4+1.4M\odot$ BNS systems. It differs from the detector horizon, the maximum detectable luminosity distance for BNS systems, which is the distance where a face-on $1.4+1.4M\odot$ BNS could produce a SNR of 8 in a single detector.

tion [148, 149, 150], estimate the Hubble constant [151, 152, 153, 154], examine gravity theories [25, 155, 156, 157, 158], investigate dark matter [159, 160], and more. While other target sources such as continuous waves and stochastic backgrounds have not been confidently detected, null results provide constraints on relevant physical models [161, 162, 163].

Looking ahead, LIGO-India is anticipated to start observation in the latter part of this decade [164], in the meanwhile LIGO, Virgo, and KAGRA will be approaching their design sensitivities. These detectors, as the successors of the first generation ran in the 2000s [165], are called the Second-Generation (2G) or 2.5G detectors. Several Third-Generation (3G) detectors are proposed to push the sensitivity to the theoretical limit of the ground-based interferometers, including Einstein Telescope (ET) [166], a 10km-long Europe-based triangular[11] interferometer and Cosmic Explorer (CE) [168, 169], a 20/40km-long US-based L-shaped interferometer. These are expected to be operational in the 2030s. Their ambitious design sensitivities, offering broader frequency bands and lower noise levels, will enable the detection of a significant fraction of CBCs in the universe and allow for precise tests of fundamental theories with unprecedented precision [170, 171]. Fig. 1.7 shows the design sensitivities of current and future ground-based detectors. The 3G detectors aim to achieve sensitivity improvements of 1-2 orders of magnitude and extend their sensitive bands to lower frequencies. Additionally, a high-frequency detector has been theoretically proposed with a novel interferometer configuration [172].

1.3.4 Other types of detectors

Limited by the seismic noise at low frequencies, ground-based detectors can not detect GWs below 1Hz, even with the advanced 3G detectors. To capture GW sources at lower frequency bands, several other types of GW detectors are proposed or constructed. I will provide a brief overview of these detectors. I omit the discussion of resonant bars as they are no longer actively investigated at the time of writing this thesis.

Space-based interferometers

Sending interferometers to the space is a direct way of avoiding seismic noises, and longer arms in the space are ideal for low-frequency GW detection. Laser Interferometer Space Antenna (LISA) [175] is the first proposed space-based GW detector. LISA consists of three spacecraft carrying lasers and test masses orbiting around each other with separations of 2.5×10^6 km. The plane of the detectors is tilted by 60° to the ecliptic, and the center of mass has an Earth-trailing heliocentric orbit and is approximately 60 million km from the Earth. LISA is sensitive in the frequency band of $10^{-4} - 1$ Hz, targeting SMBBHs, EMRIs, stochastic GWs, and stellar mass binaries. The European Space Agency (ESA)

[11] The triangular shape is the initial design. Recently, there have been discussions about triangular-shaped ET vs. two L-shaped ET at different sites, e.g. [167].

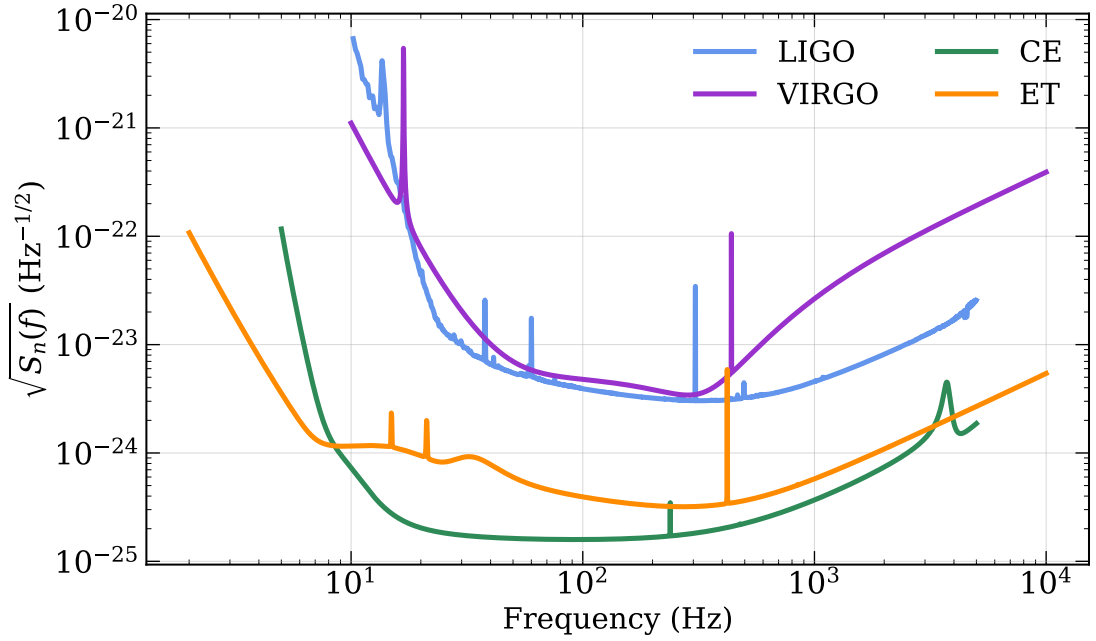


Figure 1.7: Design sensitivity of ground-based GW detectors. Sources: LIGO and Virgo: [173]; ET and CE: [174].

launched the experimental satellite LISA Pathfinder [176] in December 2015 to validate the feasibility of maintaining two spacecraft's relative position at a required precision. The experiment was successful and ESA formally adopted the LISA mission in January 2024. LISA is expected to be launched in the 2030s.

In addition to LISA, China has proposed two space-based detectors, TianQin [177] and Taiji [178] operating at similar frequency bands as LISA. Both projects feature triangular constellations. Taiji adopts a LISA-like orbit but is designed as Earth-leading, with an armlength of 3×10^6 km. TianQin, on the other hand, follows a geocentric orbit, where the normal of the detector plane points towards the reference source RX J0806.3+1527. TianQin's armlength is 1.7×10^5 km. Both of the projects are aiming to launch the detectors in the 2030s and they have launched their experimental satellites TianQin-1 [179] and Taiji-1[180].

Japan has proposed a deci-hertz space-based detector DECI-hertz Gravitational-wave Observatory (DECIGO) [181] to fill the gap between ground-based detectors and LISA-like detectors. DECIGO consists of three spacecraft in a triangular configuration on the geocentric orbit, forming three Fabry-Perot cavities with lengths around 1000 km. Putting these together, interferometer-based GW detectors will cover the GW spectrum from $\mathcal{O}(10^{-4})$ Hz to $\mathcal{O}(10^3)$ Hz.

Pulsar timing arrays

The proposal to use pulsars to detect GWs was first proposed by Sazhin [182] and Detweiler [183] and was later extended by Hellings and Downs [184]. Galactic-scale-long wavelength GWs ($10^{-9} - 10^{-6}$ Hz) could change the arrival times of radio pulses from pulsars. For a set of galactic pulsars, the changes in their arrival times caused by GWs will show correlations that depend on their angular separations, while that from noises or errors will not. By analyzing the arrival times correlation, it is expected to reconstruct the GW background on the galactic scale. The potential sources are SMBBHs and low-frequency stochastic background.

Pulsar Timing Array (PTA) is a collection of galactic millisecond pulsars monitored for GW detection. Millisecond pulsars are highly stable, making them ideal candidates for pulsar timing. Achieving significant measurements typically requires observing a sufficient number of pulsars over several years. After accumulating data for 10 years, the PTA community, including North American Nanohertz Observatory for Gravitational Waves (NANOGrav), International Pulsar Timing Array (IPTA), European Pulsar Timing Array (EPTA), Chinese Pulsar Timing Array (CPTA), and Parkes Pulsar Timing Array (PPTA), reported a confident detection of stochastic GW background in the nanohertz band [84, 85, 86, 87], opened a new window on the GW spectrum.

Other designs

In addition to the resonant bar, laser interferometer and pulsar timing, a number of other designs of GW detectors have been proposed. For example, atom interferometry is proposed for GW detection for the mHz to 10Hz band [185], and it was later extended to the space atom interferometry [186] and neutron interferometry [187]. Another approach is to use the natural objects as giant resonant bars, such as the Lunar Gravitational-wave Antenna [188], which sends inertial sensors to the Moon to monitor the response of the Moon to GWs at the mHz to 1Hz band. A similar approach has been applied on the Earth [189]. Some other methods can be used to put constraints on GWs, including CMB polarization [190], astrometry [191], and Doppler ranging of spacecraft [192].

1.4 Waveform modeling for compact binary coalescences

1.4.1 Parametrizing compact binaries

CBC sources are described by a set of parameters, which can be classified into two classes: *intrinsic* parameters and *extrinsic* parameters. The former describes the intrinsic prop-

erties of the sources and includes masses (m_1, m_2) and three-dimensional spins $(\mathbf{S_1}, \mathbf{S_2})$ of the two components, which adds up to 8 parameters. If the binary contains neutron stars, each neutron star component needs an additional parameter Λ for the tidal deformation effects (which is zero for black holes) [136], which is defined as

$$\Lambda = \frac{2k_2}{3}\left(\frac{R}{Gm}\right)^5, \tag{1.139}$$

where k_2 is the dimensionless $\ell = 2$ Love number that describes how neutron stars response to external tidal forces. R is the neutron star radius. If the orbit is eccentric, two more parameters, orbital eccentricity and eccentric anomaly are required to parametrize the orbit [110]. Extrinsic parameters do not change the physics of the source but could affect our observation. For ground-based detectors, this includes the inclination angle ι, 3D sky localizations (for example, right ascension α, declination δ, and luminosity distance d_L), polarization angle ψ, coalescence phase ϕ_c and coalescence time t_c, totaling 7 parameters. The luminosity distance d_L has a one-to-one relation with the redshift z given a cosmological model, and it is worth mentioning that the *observed* masses (often referred to as detector frame masses) are related to redshift z because GWs are redshifted by a factor $(1+z)$ during propagation at cosmological distance [6]. Since the GW frequency is inversely proportional to the mass of the system, the source frame mass and the detector frame mass are also related by the factor $(1+z)$:

$$m_{\text{detector}} = (1+z)m_{\text{source}}. \tag{1.140}$$

Mathematically, this can be derived by converting all redshift-related terms in the GW waveform Eq. 1.84 to the observables in the detector frame, including the conversions between observer clock τ_{detector} and source frame clock τ_{source}, and between the coordinate distance r and the luminosity distance d_L. Redshift factor can then be totally absorbed into the chirp mass by $\mathcal{M}_{\text{detector}} = (1+z)\mathcal{M}_{\text{source}}$, and the amplitude of GW scales down over the luminosity distance d_L instead of the coordinate distance r. Detailed calculation can be found in Ref. [6].

There can be multiple equivalent parametrizations for some parameters. The choice of parametrization can significantly impact data analysis, as certain combinations may help to avoid degeneracies, making it easier to constrain parameters effectively. For example, instead of using component masses (m_1, m_2), we can use $(\mathcal{M}, q)$, where $\mathcal{M}$ is the chirp mass defined in Eq. 1.76 and q is the mass ratio m_2/m_1[12]. Chirp mass dominates the inspiral waveform and the combination $(\mathcal{M}, q)$ has less degeneracy than (m_1, m_2). Similarly, the

[12]We usually use m_1 for the primary (heavier) component and m_2 for the secondary, therefore $0 < q \leq 1$. This is the conventional definition in data analysis as q is in a finite interval. In the waveform modeling community, the convention is $q = m_1/m_2 \geq 1$

component spins are naturally parametrized by their six spatial components (often referred to as the *radiation frame*), however, a better parametrization for data analysis is using spin magnitudes and tilt angles $(a_1, a_2, \theta_1, \theta_2, \varphi_{12}, \varphi_{JL})$ [193](often referred to as the *system frame*), which are defined as follows: a_i is dimensionless spin magnitude

$$a_i = \left| \frac{\mathbf{S_i}}{G m_i^2} \right| \tag{1.141}$$

which ranges from 0 to 1; the total angular momentum of the binary system $\mathbf{J}$ is the combination of the orbital angular momentum $\mathbf{L}$ (which is perpendicular to the orbit plane (x, y)) and the component spin angular momentum:

$$\mathbf{J} = \mathbf{L} + \mathbf{S_1} + \mathbf{S_2}. \tag{1.142}$$

An illustration of these vectors and angles is shown in Fig. 1.8. The component spins may not be aligned with the normal of the orbit, and their tilt angles to $\mathbf{L}$ are defined as $\theta_{1,2}$. In this case, the angle between $\mathbf{J}$ and $\mathbf{L}$ is φ_{JL}; it describes how much the total angular momentum deviates from the orbital normal. The angle between the observer direction $\hat{n}$ and $\mathbf{J}$ is θ_{JN}, between $\hat{n}$ and $\mathbf{L}$ is the inclination angle ι, and only one of them is enough to describe the geometry. Since $\mathbf{L}$ and $\mathbf{J}$ are misaligned, $\mathbf{L}$ will precess around $\mathbf{J}$ [194], it is therefore more convenient to use θ_{JN} to replace ι in the precessing binaries. φ_{12} is the azimuthal angle of $\hat{S}_2 - \hat{S}_1$ measured relative to $\hat{L}$. The *hat* ($\hat{}$) denotes unit vectors. In practice, these angles evolve during the binary evolution [195], so we need to assign a *reference frequency* $f_{\rm ref}$ to specify at which frequency are the angles defined.

The dominant spin parameters is the effective spin $\chi_{\rm eff}$, a mass-weighted combination along the $\mathbf{L}$ direction [196, 197, 198][13]:

$$\chi_{\rm eff} = \frac{a_1 \cos\theta_1 + q a_2 \cos\theta_2}{1+q}. \tag{1.143}$$

The effective precession spin $\chi_{\rm P}$ leads the precession effects and it takes the form [201, 202]

$$\chi_{\rm P} = \max\left(a_1 \sin\theta_1, \frac{3+4q}{4+3q} q a_2 \sin\theta_2\right). \tag{1.144}$$

$\chi_{\rm eff}$ and $\chi_{\rm P}$ are often used to quantify the spin and precession of binary systems [140, 203, 204].

Similarly, tidal parameters Λ_i in BNS systems are sometimes reparametrized as $\tilde{\Lambda}$ and

[13] Closed-form waveform formulae including spin parameters can be found in Refs. [199, 200], while the construction and use of the $\chi_{\rm eff}$ can be found in Refs. [196, 197, 198],

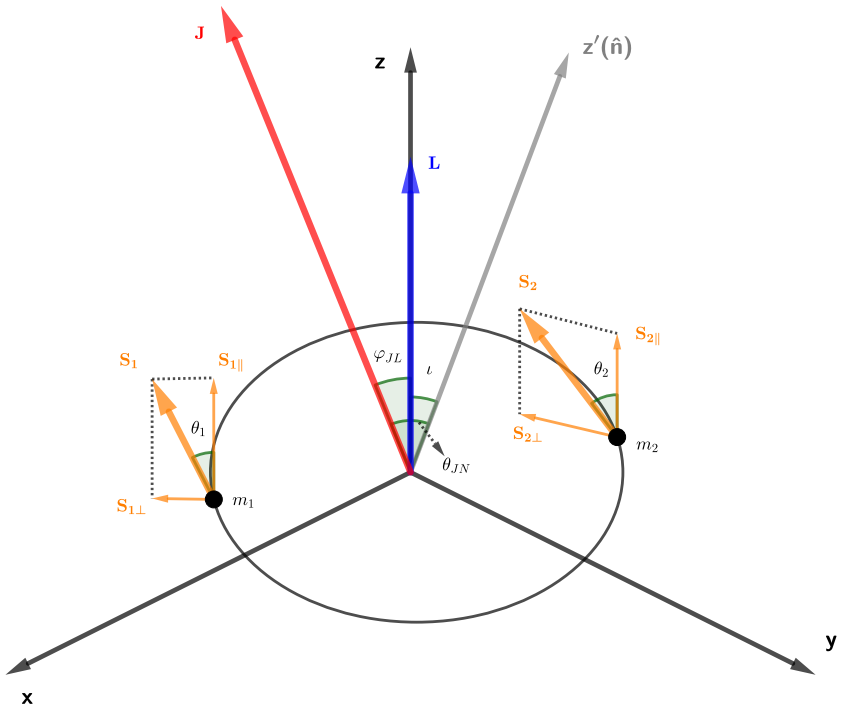


Figure 1.8: Illustration of spin angles. **J** is the total angular momentum of the binary system, **L** is the orbital angular momentum, and $\mathbf{S_1}$ and $\mathbf{S_2}$ are spin angular momentum. (x, y, z) is the radiation frame in which spin components $(S_{1x}, S_{1y}, S_{1z}, S_{2x}, S_{2y}, S_{2z})$ are defined. z' points to the observer. Orbit phase is not shown in this figure.

$\delta\tilde{\Lambda}$ [205], where

$$\tilde{\Lambda} = \frac{16}{13}\frac{(m_1+12m_2)\,m_1^4\Lambda_1 + (m_2+12m_1)\,m_2^4\Lambda_2}{(m_1+m_2)^5}. \tag{1.145}$$

$$\delta\tilde{\Lambda} = \frac{(1319m_1^2 - 7996m_1m_2 - 11005m_2^2)m_1^4\Lambda_1 - (1319m_2^2 - 7996m_1m_2 - 11005m_1^2)m_2^4\Lambda_2}{1319(m_1+m_2)^6}. \tag{1.146}$$

1.4.2 Waveform modeling

We derived the inspiral waveform for compact binaries at the leading order in Sec. 1.1.5. However, robust data analysis demands higher accuracy and more physics to be included in the waveform models. I will briefly introduce waveform modeling for CBC sources in this section.

Post-Newtonian inspiral waveform

When the binary is at large separation and low speed, Post-Newtonian (PN) approximation can be used to iteratively expand the waveform to higher orders in terms of $\mathcal{O}\left((v/c)^n\right)$. The study on PN started in the early days of GR and remains active today. A good review can be found in Ref. [206]. The order of $\mathcal{O}\left((v/c)^N\right)$ is called the $N/2$-PN order. For example, to the 3.5PN order, the GW phase of a non-spinning binary under SPA (Eq. 1.95) is modified to [207]

$$\begin{aligned}\psi_{\rm 3.5PN}(f) = 2\pi f t_c - \phi_c - \frac{\pi}{4} + \frac{3}{128\eta v^5}\Bigg[1 &+ \frac{20}{9}\left(\frac{743}{336}+\frac{11}{4}\eta\right)v^2 \\ &- 16\pi v^3 + 10\left(\frac{3058673}{1016064}+\frac{5429}{1008}\eta+\frac{617}{144}\eta^2\right)v^4 \\ &+ \pi\left(\frac{38645}{756}-\frac{65}{9}\eta\right)\left\{1+3\ln\left(\frac{v}{v_{\rm ISCO}}\right)\right\}v^5 \\ &+ \left\{\frac{11583231236531}{4694215680} - \frac{640}{3}\pi^2 - \frac{6848\gamma}{21} - \frac{6848}{21}\ln(4v)\right. \\ &\quad \left. + \left(-\frac{15737765635}{3048192}+\frac{2255\pi^2}{12}\right)\eta + \frac{76055}{1728}\eta^2 - \frac{127825}{1296}\eta^3\right\}v^6 \\ &+\pi\left(\frac{77096675}{254016}+\frac{378515}{1512}\eta-\frac{74045}{756}\eta^2\right)v^7\Bigg],\end{aligned} \tag{1.147}$$

where $\eta = m_1m_2/(m_1+m_2)^2$ is the *symmetric mass ratio*, $m = m_1+m_2$ is the total mass, $v = (\pi G m f)^{1/3}$ is the characteristic velocity in the binary. $v_{\rm ISCO}$ is the characteristic velocity at the Innermost Stable Circular Orbit (ISCO), and $\gamma = 0.577216\ldots$ is the Euler constant. The phase Eq. 1.95 corresponds to the first term $1(=v^0)$ in the square brackets, i.e., it is the 0PN order equation.

PN waveforms are analytic and therefore fast to evaluate. It also provides a clear

picture of the physics and brings insights to data analysis. There is no minus PN term in Eq. 1.147 - it could be contributed from the forbidden modes in GR, for example, the dipole radiation would appear in the -1PN order if it existed. 0.5PN naturally vanishes in GR. The 1PN term introduces mass ratio so that we can measure both component masses instead of only chirp mass. The leading-order spin effects (spin-orbit coupling) first appear in the 1.5PN order as a mass-weighted average of component spins (roughly as the form of $\chi_{\rm eff}$) [199, 200, 208]. This brings degeneracy between mass ratio and $\chi_{\rm eff}$, but introducing higher-order spin terms could relieve the degeneracy. The perpendicular components of the spins first contribute to the phase at the 2PN term (spin-spin coupling) [208, 209]. The frequency dependence is canceled at the 2.5PN, which introduces some slight degeneracy between the mass ratio and coalescence phase. The tidal parameters for BNS systems first appear in the 5PN order, implying that the tidal effect is relatively weak. However, tidal effects are detectable since their modifications on GW phase can accumulate.

Numerical relativity simulation

PN waveforms maintain good accuracy in the inspiral stage, but when the compact binary stars approach each other toward merger, the PN expansion loses validity and should be taken over by Numerical Relativity (NR) simulations [206], which numerically integrate the Einstein field equations, evolving the system from an initial state to a final state [210, 211]. Due to the complex nature of the Einstein field equations (gauge freedoms, coordinate singularities, initial conditions, boundary conditions and so forth [212]) and the limit of computational resources, NR simulation for BBH merger was not possible until the 2000s [213, 214]. Nowadays NR simulation could provide the most accurate GW waveforms of CBCs (including neutron stars [215, 216]), and there are several NR waveform catalogs available [217, 218, 219]. However, NR simulations are extremely computationally expensive, typically requiring weeks to months to generate short GW waveforms that span from the late inspiral to the final merger stages. Despite their accuracy, NR catalogs contain only hundreds or thousands of waveforms, which is relatively small given the multi-dimensional parameter space. Because of these limitations, NR waveforms are usually used to calibrate other waveform models, as a validation of other waveforms, or as the injected signal. Direct usage of NR waveforms in data analysis is not common, but possible, see e.g. Ref. [220].

It is worth mentioning that unlike the source momentum expansion (which leads to the quadruple formula Eq. 1.46) or the PN expansion, the *multipole expansion* [221] is often used NR simulations and many other waveform models. Discussions of matching these expansions can be found in Refs. [6, 206, 222]. In multipole expansion, plus and

cross polarizations are combined as a complex number and expanded:

$$h(t) = h_+(t) - ih_\times(t) = \sum_{\ell=2}^{+\infty} \sum_{m=-\ell}^{\ell} {}_{-2}Y_{\ell m}\left(\varphi_0, \iota\right) h_{\ell m}(t), \tag{1.148}$$

where (φ_0, ι) are relevant to the phase and inclination angle that locates the observer in the source frame. ${}_{-2}Y_{\ell m}$ is the *-2-spin-weighted spherical harmonics* [221, 223]. GW waveforms are calculated by solving the $h_{\ell m}(t)$, in which $(\ell = 2,\ m = \pm 2)$ are the dominant modes of GWs and others are often referred to as *higher modes* or *higher multipoles*. This decomposition is related to the outgoing transverse Weyl scalar Ψ_4, which contains all information about GWs [224] and is the derivative of GW strain [225]

$$\Psi_4 = \ddot{h}_+(t) - i\ddot{h}_\times(t). \tag{1.149}$$

Quasi-Normal Modes

Numerical relativity has shown that the remnant of a BBH merger is a perturbed Kerr black hole, which has been systematically investigated both theoretically and numerically [44, 226]. The perturbed black hole will produce damping GWs due to Quasi-Normal Modes (QNMs) of vibration; this stage is called ringdown. The ringdown GW can be decomposed into spin-weighted *spheroidal* harmonics with angular indices (ℓ, m) [227, 228, 229], which can be further written as an expansion in spin-weighted *spherical* harmonics used in NR [223, 229]. For each (ℓ, m), there exists another integer index n to characterize the waveform: $n = 0$ corresponds to the longest-lived dominant mode, while $n \geq 1$ modes are the *overtones* [230]. Generally, the ringdown waveform in the time domain can be written as the superposition of damped sinusoids [228, 231]

$$h_{\ell m}(t) = \sum_{n=0} C_{\ell mn} e^{-i\omega_{\ell mn}(t-t_0)} \quad t \geq t_0, \tag{1.150}$$

where t_0 is the time when the system exits the non-linear merger phase and the ringdown stage starts [232, 233, 234]. According to the no-hair theorem of black holes [235, 236], the characteristic frequencies (spectrum) and damping durations of the ringdown are exclusively determined by mass and spin of the remnant black hole[14]. Therefore, the complex frequencies $\omega_{\ell mn}$ are determined by the mass and spin of the remnant black hole and can be predicted by perturbation theory given the remnant mass and spin [226]. The complex-valued amplitudes $C_{\ell mn}$, however, depend on the binary configuration and dynamics near the merger and need to be fitted from the whole coalescence process, where NR is supposed to step in. The Fourier transform of Eq. 1.150 is a Lorentzian function. This analytical

[14]Black holes may also have electric charge, but this is expected to be negligible for astrophysical black holes as they could quickly attract particles of the opposite charge.

result is useful in modeling frequency domain ringdown waveforms [237].

Waveform approximants

Different methods are applied to the insprial, merger, and ringdown phases of CBC waveform modeling. However, we want the waveform model to encompass all these processes for data analysis. A number of *waveform approximants* are built for this purpose. I will go through the commonly used waveform models in this section.

The Effective-One-Body (EOB) approach [238, 239, 240] maps the relativistic two-body problem onto a test particle moving in an effective external metric. The map can be built with the help of the PN expanded equations of motion of compact binaries, which are decided by the binary properties (mass, spin, tidal deformability etc). While PN expansion loses validity near the merger, the EOB approach is able to model the insprial-merger-ringdown process in a single process [241]: it assumes that the merger is short in time and broad in frequency, and builds the merger-ringdown signal by attaching the plunge signal (representing the merger) to a superposition of QNMs (representing the ringdown). This process can be calibrated against NR simulations. To date, the EOB waveform family has evolved to the fifth generation [242, 243] from its preceders [244, 245, 246, 247, 248, 249] and has been extensively used in GW data analysis [126, 127, 150]. However, since a differential equation needs to be numerically solved during waveform evaluation, the calculation time of EOB waveforms could be much longer than other analytical models.

Inspiral-Merger-Ringdown phenomenological (IMRPhenom) waveforms are another major waveform family [237, 250, 251, 252, 253, 254, 255, 256]. Instead of physically solving the system, IMRPhenom uses analytical piecewise functions to represent the three stages of CBC waveforms and hybridize them by minimizing the difference in a hybridization region. It is usually built in the frequency domain. Motivated by the PN expansion and QNMs, the phenomenological amplitude takes the form of a linear summation of powers of frequency for inspiral and merger stage, and the Lorentzian function for the ringdown stage. The phenomenological phase can also be expanded in powers of frequency [237, 251]. The phenomenological coefficients are then calibrated against NR simulations. Being analytical brings IMRPhenom waveforms great flexibility and evaluation speed, making it the most widely used waveform model in GW astronomy.

Surrogate modeling is also widely used in GW waveform modeling. Surrogate modeling aims to use fast-to-evaluate functions such as polynomials, splines, or even neural networks to approximate the waveform functions. For example, Reduced-Order Modeling (ROM) [257, 258] is a powerful algorithm to build surrogate models for slow waveforms. It first constructs a set of linear bases (principal components) for the waveform space using Singular Value Decomposition (SVD) or greedy algorithms to choose the representative physical parameters, and then uses empirical interpolation to determine a reduced array

of time or frequency to evaluate the waveform, and uses surrogate functions (polynomials etc) to approximate waveforms at those selected times or frequencies. Finally, these approximations can be extended to the entire parameter space at all times/frequencies with a controllable error. ROM has been applied to EOB families to build fast frequency domain EOB waveforms [258, 259, 260]. It is also applied to NR simulations to build Numerical Relativity Surrogate (NRSur) family [261, 262, 263, 264], which is the most accurate waveform family since it is directly built on NR. Nevertheless, NRSur waveforms are shorter in time and are only available in the limited region in parameter space due to the same limitations of NR. It should be noted that ROM suffers from the curse of dimensionality, therefore ROM models are often limited to non-precessing waveforms or reduced parameter ranges. Several attempts have been made to develop novel surrogate modeling methods. To name a few, Ref. [265] employed neural networks to replace traditional analytical surrogate functions and enabled surrogate modeling for precessing EOB waveforms. Refs. [266, 267, 268, 269] use Gaussian progress regression to build BBH models that can give the waveform uncertainty while evaluating the waveform.

1.4.3 Waveform benchmarks

From the angle of data analysis, there are several qualities of waveform models we need to particularly pay attention to, as they are the key to successful (correct and feasible) data analysis.

Accuracy

First, all waveform models are approximate solutions to the Einstein field equations, so their *accuracy* needs to be carefully assessed. *Mismatch* compared to the true waveform is the widely-used measurement of the accuracy for waveform approximants. The mismatch between two waveforms h_1 and h_2 is defined as

$$\mathrm{MM} = 1 - \max_{t,\phi} \frac{(h_1|h_2)}{\sqrt{(h_1|h_1)(h_2|h_2)}}, \tag{1.151}$$

where $(\dots|\dots)$ is the inner product Eq. 1.137. The mismatch ranges from 0 to 1, and 0 indicates that the two waveforms are identical (or differ by a constant factor). However, due to the different parameter conventions, h_1 and h_2 may differ from each other by unphysical time and phase shift. Therefore, it is often necessary to apply a time shift and a phase shift to h_2 for mismatch calculation, i.e., $h_2 \to h_2 e^{-2\pi i f \delta t - i\delta\phi}$[15], and search the $(\delta t, \delta\phi)$ that minimizes the mismatch. The search can be performed with an inverse Fourier transform instead of a grid-based search [270]. We want to minimize

[15] From now on, I will omit the *tilde* for frequency domain waveforms h_i.

$(h_1 - h_2 e^{-2\pi i f \delta t - i\delta\phi} | h_1 - h_2 e^{-2\pi i f \delta t - i\delta\phi})$:

$$\begin{aligned}
&(h_1 - h_2 e^{-2\pi i f \delta t - i\delta\phi} | h_1 - h_2 e^{-2\pi i f \delta t - i\delta\phi}) \\
&= 2\int_{-\infty}^{+\infty} \frac{(h_1^* - h_2^* e^{2\pi i f \delta t + i\delta\phi})(h_1 - h_2 e^{-2\pi i f \delta t - i\delta\phi})}{S_n(f)} df \\
&= 2\int_{-\infty}^{+\infty} \frac{h_1^* h_1 + h_2^* h_2 - h_1 h_2^* e^{2\pi i f \delta t + i\delta\phi} - h_1^* h_2 e^{-2\pi i f \delta t - i\delta\phi}}{S_n(f)} df \\
&= ||h_1||^2 + ||h_2||^2 - 4\mathrm{Re}\left[\int_{-\infty}^{+\infty} \frac{h_1 h_2^* e^{2\pi i f \delta t} e^{i\delta\phi}}{S_n(f)} df\right] \\
&= ||h_1||^2 + ||h_2||^2 - 4\mathrm{Re}\left[e^{i\delta\phi}\int_{-\infty}^{+\infty} \frac{h_1(f) h_2^*(f)}{S_n(f)} e^{2\pi i f \delta t} df\right],
\end{aligned} \tag{1.152}$$

where

$$||h_i||^2 = (h_i | h_i) \tag{1.153}$$

is the norm of a strain. Now the problem is converted to maximizing

$$\mathrm{Re}\left[e^{i\delta\phi}\int_{-\infty}^{+\infty} \frac{h_1(f) h_2^*(f)}{S_n(f)} e^{2\pi i f \delta t} df\right] \tag{1.154}$$

Note the integral is essentially an inverse Fourier transform. Let

$$\tilde{X}(f) = \frac{h_1(f) h_2^*(f)}{S_n(f)} \tag{1.155}$$

$$X(\delta t) = \mathcal{F}^{-1}(\tilde{X}(f)) = (h_1^*(f) \mid h_2^*(f) e^{2\pi i f \delta t}) \tag{1.156}$$

Where $\mathcal{F}$ denotes the Fourier transform. We need to maximize

$$\mathrm{Re}\left[e^{i\delta\phi} X(\delta t)\right] \tag{1.157}$$

$\tilde{X}(f)$ is easy to calculate, then we can perform the inverse Fourier transform to $\tilde{X}(f)$, and find the point where $|X(\delta t)|$ reaches its maximum. The corresponding time $\delta\hat{t}$ is the optimal δt we want. Next, to maximize the real part of $e^{i\delta\phi} X(\delta\hat{t})$, $\delta\phi$ should be $\delta\hat{\phi} = -\arg X(\delta\hat{t})$, where arg denotes the argument of a complex number. This shows $\delta\hat{\phi}$ and $\delta\hat{t}$ have different tasks: $\delta\hat{t}$ is in charge of maximizing the module of $X(t)$, while $\delta\hat{\phi}$ is used to rotating $X(t)$ to the real axis so that $\mathrm{Re}X(t) = |X(t)|$.

The true waveform is unknown in practice, so waveform approximants are usually compared with NR waveforms. Depending on the complexity of the binary system, the latest IMRPhenom waveform `IMRPhenomXPHM` [254] and EOB waveform `SEOBNRv5PHM` [242] could achieve mismatch of $10^{-5} - 10^{-1}$ and a median around $10^{-3} - 10^{-2}$. It is reported that `SEOBNRv5PHM` is more accurate than `IMRPhenomXPHM` [242] but it is slower to compute. ROMs can usually achieve $< 10^{-3}$ mismatch against the original waveform they are built

upon. The latest NRSur model `NRSur7dq4` [262] outperforms IMRPhenom and EOB in accuracy, with mismatch of $10^{-6}-10^{-2}$ and a median around $10^{-4}-10^{-3}$. We should note that even NR simulations have an estimated mismatch above 10^{-7} and up to 10^{-1} [271].

Waveform models are accurate enough when they do not cause systematic errors in data analysis and louder signals need more accurate waveforms. This will be discussed in the next chapters, and I will show that the current waveforms are marginally enough for current GW detections and the mismatch should be improved by $3-4$ orders of magnitude for the 3G detectors.

Speed

Waveform approximants may be precomputed before data analysis (e.g. for template bank [272]) and may also be calculated during data analysis (e.g. for parameter estimation [273]). It is clear that the latter is more demanding on the speed of waveform generation.

A typical Bayesian parameter estimation needs millions of waveform evaluations [274], thus a fast waveform is crucial. The generation time depends on the length of the waveform - it gets slower if we set a lower minimum frequency or lower masses. It also depends on the complexity of the system. For stellar-mass CBC waveforms from 20Hz, analytical waveforms (PN, IMRPhenom, ROM) can usually be generated within 100ms. EOB waveforms need to solve an ordinary differential equation during waveform generation, so their generation time is much longer, typically at $\mathcal{O}(1)$s, although the latest EOB waveform has reported the improvements of time cost to $\mathcal{O}(100)$ms [242]. In addition to the traditional waveform generation platforms which are on CPU and based on `C`, GPU acceleration of waveform computation has been proposed and achieved $\mathcal{O}(10)$ times speed improvement on large GPUs [275].

Higher-order effects

Compact objects naturally have spin. When the spin angular momentum is not aligned with the orbital angular momentum, the binary system will precess. If the binary is assembled dynamically, it may have an eccentric orbit instead of a quasi-circular one, although the orbit would gradually circularize due to the energy loss. Additionally, if the binary is edge-on to us and has unequal masses, we may observe the subdominant multipoles. Fig. 1.9 shows comparisons between waveforms with precession, eccentricity, higher modes and "vanilla" aligned-spin waveforms. All of them induce amplitude modulations, but at different magnitudes and time scales [276]. These modulations are hard to distinguish if the signal is not long or loud enough [277].

These phenomena are intriguing but hard to model, and missing them may lead to a biased data analysis [278, 279]. Therefore, they have received considerable attention in

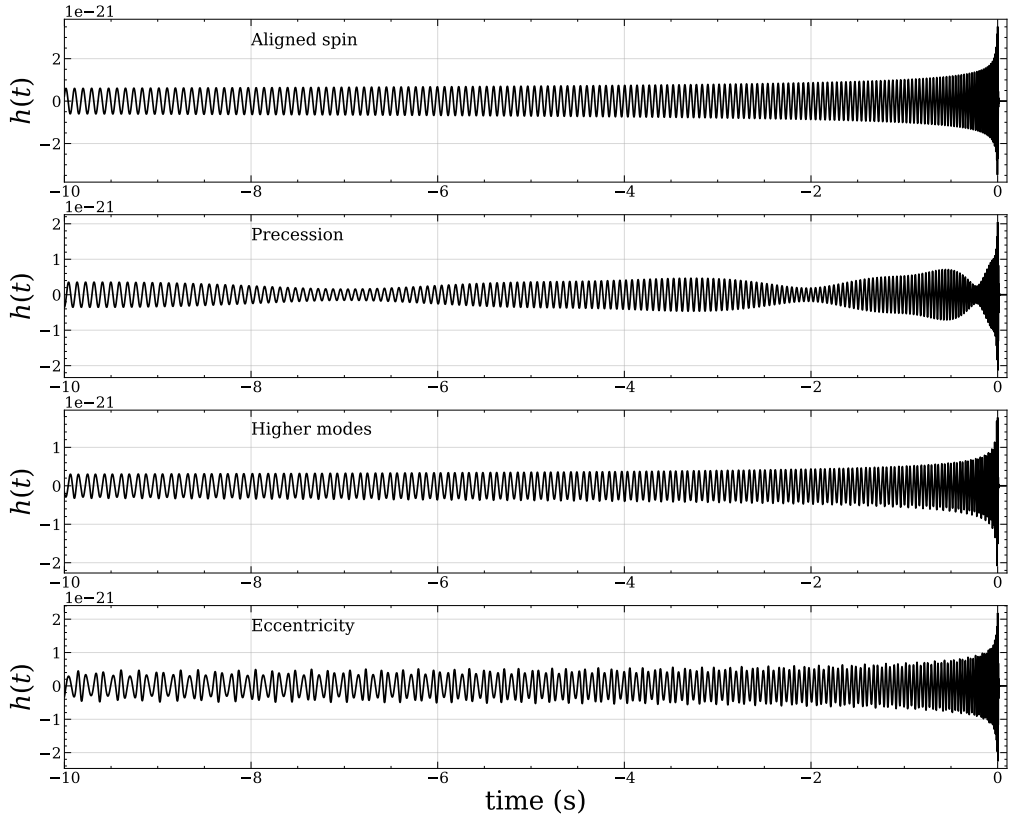


Figure 1.9: Illustrations of different higher-order effects of GWs for $50 + 10M_{\odot}$ binaries. The plus polarization is plotted. The aligned spin waveform model is `IMRPhenomD`, the precessing model is `IMRPhenomPv2`, the higher modes model is `SEOBNRv4HM`, and the eccentric model is `SEOBNRE`. The choice of waveform models and inclination angles, spins, and eccentricity parameters ensure that the effects are visible and only the target effect is visible.

the waveform modeling community. With progresses in NR simulations [280, 281], waveform approximants with higher modes for precessing systems (e.g. `IMRPhenomXPHM` [254], `SEOBNRv5PHM` [242] and `NRSur7dq4` [262]) and for eccentric systems (e.g. `SEOBNRv4EHM` [249], `SEOBNREHM` [282], `NRSur2dq1Ecc` [264]) have been developed. However, as NR simulations are sparse in the spin and eccentricity space, these models are not optimally calibrated and become less accurate when precession or eccentricity effects are strong. Moreover, it is challenging to include both precession and eccentricity effects in a single waveform model, so the aforementioned waveform approximants typically incorporate only one of these effects. A recent attempt at including both can be found in Ref. [283].

1.5 Summary

Focusing on CBCs and the ground-based detectors, I introduced the theoretical background of GW astronomy in this chapter. I started with GR and linearized gravity and demonstrated the polarization of GWs and the quadruple formula of GW generation, and then derived the chirp waveform of inspiralling binaries in Sec. 1.1. The CBC waveform is further discussed in Sec.1.4. On the observation side, I introduced various types of astrophysical sources of GWs in Sec. 1.2 and described the configurations, response, and noise characteristics of GW detectors in Sec. 1.3. These materials aim to provide a context for the research projects in this thesis.

Chapter 2

Data analysis for compact binary coalescences

Data analysis is the bridge connecting physical theories and GW observations. After formulating sophisticated physical theories and designing advanced GW detectors, the collected data must be accurately and efficiently interpreted. This allows us to identify signals within noisy data and analyze the properties of their sources. In this chapter, I will introduce the detection and parameter estimation of GWs, including both traditional statistical methods in Sec. 2.1 and 2.2 and novel machine learning methods in Sec. 2.3. Additionally, I will address the challenges in current and future GW data analysis in Sec. 2.4. A summary of this chapter and the goal of this thesis are given in Sec. 2.5.

2.1 Detecting signals from noisy data

The GW data d consists of signal and noise, and *detecting* the signal in the noisy data is the first step of all subsequent physics-related analyses. In this section, I will introduce the basic matched filtering search used in LVK detection and briefly introduce current detection pipelines.

2.1.1 Matched filtering

Following Brown [284], we show why and how we use matched filtering for searching GW signals.

Why matched filter: a probabilistic view

Identifying possible signals present in the data can be viewed as a statistical hypothesis test problem. The null hypothesis is $\mathcal{H}_0$: There is no signal present in the data, i.e.,

$$d = n. \tag{2.1}$$

The alternative hypothesis is $\mathcal{H}_1$: There is a signal h present in the data, i.e.,

$$d = h + n. \tag{2.2}$$

When choosing between these two hypotheses, we may encounter two types of errors. The first type is *missed detection*, where $\mathcal{H}_1$ is true but we mistakenly choose $\mathcal{H}_0$. The second is *false alarm*, where $\mathcal{H}_1$ is false but we incorrectly choose it. The latter error is more detrimental, and we typically aim to avoid it in hypothesis testing.

The Neyman-Pearson lemma [285] shows that the likelihood ratio test is the optimal test for binary hypothesis since it minimizes the false dismissal probability for a given false alarm probability. From the noise likelihood Eq. 1.136, the likelihood ratio between $\mathcal{H}_1$ and $\mathcal{H}_0$ can be written as

$$\mathcal{L} = \frac{p(d|\mathcal{H}_1)}{p(d|\mathcal{H}_0)} = \frac{e^{-\frac{1}{2}(d-h|d-h)}}{e^{-\frac{1}{2}(d|d)}} = e^{(d|h)} e^{-\frac{1}{2}(h|h)}. \tag{2.3}$$

The $(h|h)$ is a constant given a waveform h and PSD. This means that, from the perspective of the likelihood ratio test, $(d|h)$ is the only relevant term to decide whether h is present in d and the likelihood ratio increases monotonically with this term. This implies that $(d|h)$ is relevant to the optimal detection statistic: if it exceeds a certain threshold, we may reject $\mathcal{H}_0$ and assert the detection of a signal. The inner product term $(d|h)$ is commonly referred to as the *matched filter*, as it effectively represents a cross-correlation between the

expected signal and the data, quantifying their similarity.

Signal-to-noise ratio: a signal processing view

We have demonstrated that matched filtering serves as the optimal detection statistic that minimizes false alarms from a probabilistic perspective. Alternatively, we can think from the angle of signal processing. In fact, the matched filter is the filter that maximizes the Signal-to-Noise Ratio (SNR), a widely used quantity in astronomy. Let K denote the filter function that yields the filter output (cross-correlation) z:

$$\begin{aligned} z &= \int_{-\infty}^{\infty} d(t)K^*(t)dt = \int_{-\infty}^{\infty} \tilde{d}(f)\tilde{K}^*(f)df \\ &= \int_{-\infty}^{\infty} \tilde{h}(f)\tilde{K}^*(f)df + \int_{-\infty}^{\infty} \tilde{n}(f)\tilde{K}^*(f)df \\ &= H + N. \end{aligned} \tag{2.4}$$

where H is fixed for a fixed waveform and measures the strength of the signal, while N is a random variable with zero mean and measures the strength of the noise. $\langle N^2 \rangle$ is the ensemble average and can be calculated:

$$\begin{aligned} \langle N^2 \rangle &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{K}^*(f)\tilde{K}^*\left(f'\right)\langle \tilde{n}(f)\tilde{n}\left(f'\right)\rangle df df' \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tilde{K}^*(f)\tilde{K}^*\left(f'\right)\frac{1}{2}S_n(|f|)\delta\left(f-f'\right)df df' \\ &= \frac{1}{2}\int_{-\infty}^{\infty} |\tilde{K}(f)|^2 S_n(|f|)df, \end{aligned} \tag{2.5}$$

We want the filter K to maximize $H^2/\langle N^2 \rangle$ so that the signal stands out, and the ratio is given by

$$\begin{aligned} \frac{H^2}{\langle N^2 \rangle} &= \frac{2\left|\int_{-\infty}^{\infty} \tilde{h}(f)\tilde{K}^*(f)df\right|^2}{\int_{-\infty}^{\infty} |\tilde{K}(f)|^2 S_n(|f|)df} \\ &= \frac{2\left|\int_{-\infty}^{\infty} \left[\tilde{h}(f)/\sqrt{S_n(|\,f\,|)}\right]\left[\tilde{K}^*(f)\sqrt{S_n(|\,f\,|)}\right]df\right|^2}{\int_{-\infty}^{\infty} |\tilde{K}(f)|^2 S_n(|f|)df} \\ &\leq 2\int_{-\infty}^{\infty} \frac{|\tilde{h}(f)|^2}{S_n(|\,f\,|)}df, \end{aligned} \tag{2.6}$$

where in the last line we have used the Cauchy-Schwarz inequality

$$\left|\int_{-\infty}^{\infty} A(f)B(f)df\right|^2 \leq \int_{-\infty}^{\infty} |A(f)|^2 df \int_{-\infty}^{\infty} |B(f)|^2 df, \tag{2.7}$$

where $A(f) = \tilde{h}(f)/\sqrt{S_n(|f|)}$ and $B(f) = \tilde{K}^*(f)\sqrt{S_n(|f|)}$. The two sides are equal if and only if $A(f) = C \cdot B(f)$, and here C can be any non-zero constant. Let $C = 2$, we have

$$\tilde{K}^*(f) = 2\frac{\tilde{h}^*(f)}{S_n(f)}. \tag{2.8}$$

Substituting $\tilde{K}(f)$ into the filter output z, we get [1]

$$z = 2\int_{-\infty}^{\infty} \frac{\tilde{d}(f)\tilde{h}^*(f)}{S_n(f)} df = (d|h), \tag{2.9}$$

which coincides with the optimal detection statistic constructed by the likelihood ratio test. The noise level $\langle N^2 \rangle$ becomes

$$\langle N^2 \rangle = \frac{1}{2}\int_{-\infty}^{\infty} 4\frac{|\tilde{h}(f)|^2}{S_n(|f|)} df = (h|h). \tag{2.10}$$

When the data contains no signal, the matched filtering output $z = N = (n|h)$, which is a Gaussian with zero mean because the inner product is essentially a linear summation of Gaussian noise n that has zero mean. The variance of this distribution is $(h|h)$. The *matched filtering SNR* is defined as

$$\rho = \frac{(d|h)}{\sqrt{(h|h)}}. \tag{2.11}$$

By this definition, ρ follows the standard Gaussian distribution when the data is pure noise. The *optimal SNR* is the SNR when there is no noise, i.e.

$$\rho_{\rm opt} = \sqrt{(h|h)}. \tag{2.12}$$

On the ensemble level, ρ will be a Gaussian distribution centered at $\rho_{\rm opt}$ with the variance of 1 when a signal is present.

Search over unknown parameters

The matched filter output depends on the waveform h, which further depends on source parameters $\boldsymbol{\theta}$. From the probabilistic view, the likelihood ratio should be marginalized over $\boldsymbol{\theta}$ with a prior distribution of $p(\boldsymbol{\theta})$ [286]:

$$\mathcal{L} = \int \mathcal{L}(\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}. \tag{2.13}$$

[1] This integral gives a complex number, however, its imaginary part is typically negligible because the filter (viewed from the time domain) is a cross-correlation between the real signal and a real template which results in a real number. Hence, we disregard the imaginary part and write it in the inner product form.

Some parameters can be marginalized or maximized analytically, while some can not. For those parameters that cannot be marginalized, dense samples are drawn in the parameter space to construct waveform templates (referred to as a *template bank*) and perform matched filtering for all of them [272]. The template with the highest SNR and surpassing the threshold is selected as the preferred template, and its parameters are considered to be those of the source. Current LVK detections use aligned-spin templates, which contain 4 intrinsic parameters to be searched within the template bank (two masses and two spin magnitudes). The template bank should be dense enough such that any aligned-spin signal can find a template with a mismatch of less than 0.03. This requirement results in hundreds of thousands to millions of templates in the search bank [287]. However, the aligned-spin search is less sensitive to eccentric and precessing signals, and novel methods have been proposed for these sources, e.g. [220, 288].

There are 7 extrinsic parameters to be searched: coalescence time t_c and phase ϕ_c, sky location (α, δ), polarization angle ψ, luminosity distance d_L and inclination angle ι. For the coalescence time, we can straightforwardly shift the matched filtering template and calculate SNR at each time step, resulting in an SNR timeseries [270]:

$$\begin{aligned}\tilde{h}(f) &= \int h(t) e^{-2\pi i f t} dt \\ &= \left[\int \tilde{h}(t) e^{-2\pi i f (t+t_c)} dt\right] e^{2\pi i f t_c} \\ &= \mathcal{F}\left[h\left(t - t_c\right)\right] \cdot e^{2\pi i f t_c},\end{aligned} \tag{2.14}$$

where $\mathcal{F}$ denotes Fourier transform. The matched filtering timeseries is

$$z\left(t_c\right) = 2 \int_{-\infty}^{\infty} \frac{\tilde{d}(f)\tilde{h}^*(f)}{S_n(|f|)} e^{2\pi i f t_c} df. \tag{2.15}$$

The t_c that maximizes the matched filtering output is considered as the arrival time. For the other parameters, we begin by decomposing the waveform given in Eq. 1.84 into separate components h_c and h_s, which represent the cosine and sine parts at a unit distance:

$$\begin{aligned} h_+(t) &= \tfrac{1+\cos^2\iota}{2 d_L} \bar{\mathcal{A}}(t) \cos(\Phi(t) + \phi_c) \\ h_\times(t) &= \tfrac{\cos\iota}{d_L} \bar{\mathcal{A}}(t) \sin(\Phi(t) + \phi_c), \\ h_c(t) &= \bar{\mathcal{A}}(t) \cos(\Phi(t)) \\ h_s(t) &= \bar{\mathcal{A}}(t) \sin(\Phi(t)) \end{aligned} \tag{2.16}$$

where $\bar{\mathcal{A}}(t) = \mathcal{A}(t, d_L = 1\text{Mpc})$. Define $\sigma^2 = (h_c|h_c)$, we can verify that

$$(h_c|h_c) = (h_s|h_s) = \sigma^2, \tag{2.17}$$

and

$$(h_c|h_s) = 0 \tag{2.18}$$

The detector response is then

$$\begin{aligned} h &= F_+ h_+ + F_\times h_\times \\ &= F_+ \frac{1+\cos^2\iota}{2d_L}\bar{\mathcal{A}}(t)\cos(\Phi(t)) + F_\times \frac{\cos\iota}{d_L}\bar{\mathcal{A}}(t)\sin(\Phi(t)) \\ &= \frac{(1\text{ Mpc})}{d_{\text{eff}}}\bar{\mathcal{A}}(t)\cos\left(\Phi(t)-\phi_0\right) \\ &= \frac{(1\text{ Mpc})}{d_{\text{eff}}}\left[h_c(t)\cos\phi_0 + h_s(t)\sin\phi_0\right], \end{aligned} \tag{2.19}$$

where

$$d_{\text{eff}} = \frac{d_L}{\sqrt{F_+^2\left(1+\cos^2\iota\right)^2/4 + F_\times^2(\cos\iota)^2}} \tag{2.20}$$

is the effective distance and

$$\phi_0 = \arctan\left(\frac{F_\times\cdot(2\cos\iota)}{F_+\cdot(1+\cos^2\iota)}\right) - \phi_c \tag{2.21}$$

is an unknown phase. The six remaining extrinsic parameters are simplified to the effective distance d_{eff} and an unknown phase ϕ_0. To determine ϕ_0, we perform matched filtering for h_c and h_s respectively and construct a complex output [270]:

$$(d|h_c) + i(d|h_s), \tag{2.22}$$

and the ϕ_0 should maximize its modulus. In essence, we use $(d|h_c), (d|h_s)$ as two bases on a complex plane and determine the optimal ϕ_0 that best fits these bases. The d_{eff} acts solely as an amplitude factor - it has no impact on the SNR because SNR is computed by normalized templates by definition. The d_{eff} that maximizes the likelihood ratio can be obtained analytically by taking the derivative of the log likelihood ratio (log Eq. 2.3) with respect to d_{eff}, which yields

$$\frac{1}{d_{\text{eff}}} = \frac{(d|h_c(t)\cos\phi_0 + h_s(t)\sin\phi_0)}{\sigma^2}. \tag{2.23}$$

and the maximized log likelihood ratio is $\frac{1}{2}\rho^2$. Putting these together, the matched filtering SNR is a complex timeseries in practical detection scenarios:

$$\rho(t_c) = \frac{1}{\sigma}\left[(d|h_c(t-t_c)) + i\,(d|h_s(t-t_c))\right], \tag{2.24}$$

and its modulus is the SNR typically referenced. h_c, h_s depend on intrinsic parameters

and are searched across a template bank. The optimal arrival time should maximize $|\rho(t_c)|$ and the optimal phase is determined by real and imaginary parts of the maximum of $\rho(t_c)$. The remaining extrinsic parameters are packed into $d_{\rm eff}$. Although their optimal values cannot be directly obtained through matched filtering, the combined parameter $d_{\rm eff}$ can be derived using Eq. 2.23.

For a detector network that consists of multiple detectors, we define the inner product between matrices whose elements are time (or frequency) series:

$$\mathbf{C} = (\mathbf{D}|\mathbf{B}) \Rightarrow C_{jk} = (D_{jp} \mid B_{pk}) \,, \tag{2.25}$$

the matrix inner product produces a new matrix whose elements are calculated in analogy to matrix multiplication rules. The SNR matrix is a vector of length of N, where N is the number of detectors:

$$\boldsymbol{\rho} = (\mathbf{H_c} \mid \mathbf{d}) + i(\mathbf{H_s} \mid \mathbf{d}), \tag{2.26}$$

where

$$\mathbf{H_{c,s}} = \mathrm{diag}\left(\frac{h_{c,s}}{\sigma^{(1)}}, \frac{h_{c,s}}{\sigma^{(2)}}, \ldots, \frac{h_{c,s}}{\sigma^{(N)}}\right). \tag{2.27}$$

Each element corresponds to the SNR timeseries of a detector. SNRs in different detectors should be combined to measure the strength of a signal. We define the *network SNR* as

$$\rho_{\rm net} = \sqrt{\boldsymbol{\rho}^{\boldsymbol{T}}\boldsymbol{\rho}^{*}}, \tag{2.28}$$

i.e., the square root of the quadratic sum of all single detector SNRs. Network SNR being higher than 8 is typically the primary criterion for detection, with additional criteria discussed in subsequent sections.

In the end, it is necessary to mention that this formalism requires aligned-spin quasi-circular orbit, i.e., h_+ and $h_\times$ only have a phase difference of $\pi/2$ and satisfy $(h_+|h_\times) = 0$, so that the six extrinsic parameters can be absorbed into $d_{\rm eff}$ and ϕ_0. However, this assumption does not hold for eccentric or precessing sources. Detecting eccentric and precessing signals using aligned-spin templates will result in a loss of SNR.

2.1.2 Detection pipelines

In real detection pipelines, numerous additional considerations arise. Matched filtering across millions of templates is computationally intensive, therefore SVD is often applied to reduce the template bank size [289]. Non-Gaussian artifacts in detector data, known as *glitches*, occasionally resemble real signals and can interfere with detection. χ^2 veto is used to discriminate the noise artifacts against real signal [270], as glitches and signals typically yield SNR timeseries with different distributions [290]. Furthermore, a coincidence test

between detectors is performed based on the idea that the signal should arrive at different detector sites within a narrow time window [291, 292, 293]. In some pipelines, this is done by a coherent search, which searches coalescence time together with other source parameters [290]. Subsequently, a ranking statistic is computed for all triggered templates, where a higher value indicates higher preference [292, 294]. False Alarm Rates (FARs) are then assigned to each trigger, showing the chance of such a coincidence between detectors being caused by noise [290, 293]. Additionally, the probability of astrophysical origin, $p_{\rm astro}$, for each trigger is calculated for CBC searches [295, 296]. Following the workflow, several matched-filtering-based search pipelines are currently operational for LVK detections: GstLAL [292, 297], PyCBC Live [298, 299], MBTA [291], and SPIIR [300]. An overview of these pipelines and the thresholds of detection statistics can be found in Ref. [301]. There are several third-party search pipelines on LVK data as well, e.g. the OGC [302, 303, 304, 305] and the IAS catalog [306, 307, 308, 309]. These pipelines are based on different software and template banks and have different methods for FAR and $p_{\rm astro}$. Cross-validation among these independent pipelines ensures robust scientific outcomes.

In addition to matched-filter-based searches, there are search pipelines that assume no prior knowledge of the waveform and only make use of the coherence between detectors. They are able to detect unmodelled signals and CBCs that are not well represented in the template banks. This was actually the case in the first detection GW150914 [7, 310], when lighter black holes were initially expected. However, the unmodelled searches are mainly sensitive to short-duration signals, and they are generally less sensitive to CBCs than matched filtering methods because of the lack of knowledge of waveforms [310], although it is possible to approximately reconstruct the waveform by wavelets [311, 312]. Current unmodelled search pipelines include cWB [313, 314] and oLIB [315]. They target all kinds of short-duration signals, including high-mass CBCs.

Machine learning detection has been employed since O4. I will discuss machine learning's applications in GW astronomy in Sec. 2.3.

2.2 Inferring properties of gravitational wave sources

As discussed in Sec. 1.4.1, compact binaries are described by a set of parameters. Inferring their properties is therefore a problem of estimating these parameters. In this section, I will introduce Bayesian inference, the current "standard" method of CBC PE.

2.2.1 Bayes theorem

Matched filtering can give an initial estimate of some source parameters based on the best-fit template, akin to a maximum likelihood estimate, albeit over a reduced parameter space. The Bayesian framework suits this problem better [273, 316]. We want to estimate parameter $\boldsymbol{\theta}$ based on the observation $\boldsymbol{d}$, denoted as $p(\boldsymbol{\theta} \mid \mathbf{d})$. According to Bayes' theorem, this can be computed from the *likelihood* $p(\mathbf{d} \mid \boldsymbol{\theta})$ and a *prior* distribution $p(\boldsymbol{\theta})$:

$$\underbrace{p(\boldsymbol{\theta} \mid \mathbf{d}, \mathcal{H})}_{\text{Posterior}} = \frac{\overbrace{p(\mathbf{d} \mid \boldsymbol{\theta}, \mathcal{H})}^{\text{Likelihood}} \overbrace{p(\boldsymbol{\theta} \mid \mathcal{H})}^{\text{Prior}}}{\underbrace{p(\mathbf{d} \mid \mathcal{H})}_{\text{Evidence}}}, \tag{2.29}$$

where we have inserted $\mathcal{H}$, the model that describes the data, for example, a CBC waveform model. $p(\mathbf{d} \mid \mathcal{H})$ is the *evidence* (or marginalized likelihood), showing the probability of observing the data given the model $\mathcal{H}$. It is a constant that can be omitted in parameter estimation, but plays an important role in model comparison. The prior distribution reflects our initial knowledge about the source parameters before making observations. In the absence of specific prior information for a parameter, priors can be chosen based on the nature of the parameter itself (e.g. isotropic priors for sky locations) or simply assigned as uniform. Priors based on astrophysical models could also be used. The likelihood is the probability of observing the data given the parameters and the model. It is constructed based on the noise likelihood Eq. 1.136: after subtracting the waveform from the data, the remaining residuals are assumed to be noise and should adhere to the noise model [111]. Thus,

$$p(\mathbf{d} \mid \boldsymbol{\theta}, \mathcal{H}) \propto e^{-\frac{1}{2}(\mathbf{d}^{\mathrm{T}} - \mathbf{h}^{\mathrm{T}}(\boldsymbol{\theta}) | \mathbf{d} - \mathbf{h}(\boldsymbol{\theta}))} = e^{-\frac{1}{2}\sum_{i=1}^{N}(d_i - h_i(\boldsymbol{\theta}) | d_i - h_i(\boldsymbol{\theta}))}, \tag{2.30}$$

where we have omitted the normalization factor in likelihood. N denotes the number of detectors and i denotes the ith detector. h is the waveform model in the model $\mathcal{H}$.

2.2.2 Stochastic sampling

By combining prior and likelihood and dropping the normalization constants, we can calculate the posterior probability given any parameter $\boldsymbol{\theta}$. However, due to the high dimensionality of the parameter space, it is difficult to find the optimal estimation by the analytical posterior. In practice, we employ stochastic sampling algorithms to draw samples from the posterior distribution, and the posterior samples effectively characterize the posterior distribution itself [273]. The two primary families of sampling algorithms used are Markov Chain Monte Carlo (MCMC) and nested sampling.

Markov Chain Monte Carlo

MCMC draws samples with a Markov Chain in the parameter space. For example, in the Metropolis-Hastings MCMC algorithm [317, 318], we sample $\boldsymbol{\theta}$ from the prior, and this sample "walks" in the parameter space. The next step $\boldsymbol{\theta'}$ is proposed near $\boldsymbol{\theta}$ according to a conditional probability $q(\boldsymbol{\theta'} \mid \boldsymbol{\theta})$, and the new step will be accepted with the probability

$$r\left(\boldsymbol{\theta}, \boldsymbol{\theta}'\right) = \min\left\{\frac{p\left(\boldsymbol{\theta}' \mid \mathbf{d}\right)}{p(\boldsymbol{\theta} \mid \mathbf{d})} \frac{q\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}'\right)}{q\left(\boldsymbol{\theta}' \mid \boldsymbol{\theta}\right)}, 1\right\}. \tag{2.31}$$

If the step is accepted, the walker should proceed from $\boldsymbol{\theta'}$ and repeat this process, otherwise the walker continues from $\boldsymbol{\theta}$. The trace of the walker will eventually converge to the target distribution, and posterior samples are obtained by removing the initial unconverged steps and correlated points on the chain until a sufficient number of independent samples are collected. However, the evidence Z is inaccessible in this process. It is possible to obtain the evidence Z by tempering mechanisms [273, 319]: introducing a temperature T, and sampling the following distributions for different T:

$$p(\boldsymbol{\theta} \mid \mathbf{d}, \beta) \propto p(\boldsymbol{\theta}) p(\mathbf{d} \mid \boldsymbol{\theta})^{\beta}, \tag{2.32}$$

where $\beta = 1/T$. The log evidence can be estimated by integrating out β:

$$\ln Z = \int_0^1 \langle \ln p(\mathbf{d} \mid \boldsymbol{\theta}) \rangle_{\beta} d\beta, \tag{2.33}$$

where $\langle \ln p(\mathbf{d} \mid \boldsymbol{\theta}) \rangle_{\beta}$ is the expectation value of the log likelihood for the chain with temperature $T = 1/\beta$.

MCMC becomes much slower for high dimensional parameter spaces and narrow target distributions, as the walker may struggle to efficiently navigate to the correct next step. There are many other more advanced variants of MCMC algorithms with different step proposals (e.g. Hamiltonian Monte Carlo [320]). A review of applications of MCMC algorithms can be found in Ref. [321].

Nested sampling

The nested sampling algorithm was first proposed by Skilling [322] primarily for calculating the evidence, with sampling as a by-product. The likelihood can be written as a function (L) of the prior volume X, where $dX = p(\boldsymbol{\theta}) d\boldsymbol{\theta}$. Here $X = 1$ represents the whole prior space, and smaller X represents a smaller volume in the prior space and therefore a higher likelihood. We have

$$Z = \int p(\boldsymbol{\theta}|d) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} = \int_0^1 L(X) \mathrm{d}X. \tag{2.34}$$

Nested sampling starts with a set of live points drawn from the prior space, initializing at $X = X_0 = 1$. The point with the lowest likelihood L_0 will be removed, and a new point will be sampled from the shrunk prior volume expended by the remaining live points (a MCMC or another sampling method is required to do this). X reaches a smaller value $X_1 < 1$ in this process, and the numerical integral of Z accumulates contributions numerically as $(X_0 - X_1)L_0$. Through this iterative process, the prior space progressively shrinks, X steps to lower values, and the Z gains more contributions. The termination criterion occurs when further iterations contribute negligibly to Z, indicating that little new information is gained (for example, $d \log Z < 0.1$). At this point, the prior volume has contracted to approximate the posterior distribution and Z has accumulated to its estimated value. Nested sampling suffers from high dimensionality and complex likelihood space too. There are also many variants of nested sampling algorithms (e.g. `dynesty` [323]) and a review of physical applications of nested sampling can be found in Ref. [324].

Diagnosis of posterior distributions

Stochastic sampling methods are designed to converge to the true posterior, but they may struggle to find the truth and may terminate before reaching convergence. In practical applications where the true distribution is often unknown, but results from different samplers can be compared. Consistency among these results indicates convergence of the samplers. The difference between probability distributions can be quantified by Kullback-Leibler Divergence (KLD) (relative entropy) or Jensen-Shannon Divergence (JSD).

Consider a random variable P with probability distribution $p(x)$, its *Shannon information* is defined as

$$h(x) = -\log_b(p(x)), \tag{2.35}$$

here b is the base of the logarithm. Base 2 results in units of *bits*, and base e (natural logarithm) gives *nats*. By this definition, $h(x)$ is higher in low-probability regions because we can infer more information with the occurrence of unlikely events. The *entropy* is defined as the expectation of information

$$H(P) = \int p(x)h(x)dx = -\int p(x)\log_b(p(x))dx. \tag{2.36}$$

Now consider another random variable $Q \sim q(x)$. The KLD characterizes the relative information content between P and Q. The relative entropy (KLD) from Q to P is defined as

$$D_{\mathrm{KL}}(P\|Q) = \int p(x)\log_b \frac{p(x)}{q(x)}\mathrm{d}x, \tag{2.37}$$

KLD ranges from 0 to $+\infty$; 0 means two distributions are identical. KLD is not symmetric, i.e., $D_{\mathrm{KL}}(P\|Q) \neq D_{\mathrm{KL}}(Q\|P)$, so it is not a metric. JSD is defined based on KLD and is

constructed to be symmetric and finite:

$$D_{\rm JS}(P\|Q) = \frac{1}{2}D_{\rm KL}(p\|m) + \frac{1}{2}D_{\rm KL}(q\|m), \tag{2.38}$$

where $M = (P+Q)/2 \sim m$ is the mixture of the two distributions. We have $D_{\rm JS}(P\|Q) = D_{\rm JS}(Q\|P)$ and

$$0 \leq D_{\rm JS}(P\|Q) \leq \log_b(2). \tag{2.39}$$

JSD is often used to quantify the difference between probability distributions.

Another method to measure the difference between probability distributions is the Kolmogorov-Smirnov Test (KS Test), which measures the maximum difference between the cumulative distribution of two random variables. However, in the context of GW astronomy, KS Test is often used for another diagnostic purpose: the *self-consistency test*, also known as the *P-P test*. For a set of simulations in which true parameters are known, the percentiles of the injected parameters in the estimated posterior distributions, denoted as $p_{\rm search}$, ideally should follow a uniform distribution $U(0,1)$. Deviations from this uniform distribution suggest that the estimator is shifted or that the estimated posterior distribution is either too narrow or too wide. KS Test can be performed between a collection of $p_{\rm search}$ and $U(0,1)$, which will produce a p-value describing the probability of seeing such an extreme case. If the p-value<0.05, we should reject the null hypothesis that $p_{\rm search}$ is drawn from $U(0,1)$[2] and the estimator fails the KS Test.

The self-consistency test is often visualized by the P-P plot. Given a search confidence level p, the fraction of events whose real parameters being searched within p from the estimated posteriors is p'. When plotting the relationship between p and p' – effectively the cumulative distribution of the $p_{\rm search}$ – the line should closely follow the diagonal line between $(0,0)$ and $(1,1)$. The statistical deviation from the diagonal line can be estimated by a binomial distribution $B(k;N,p)$, where N is the number of events and k is the number of events that are searched within probability p. k follows the binomial distribution $B(k;N,p)$, whose continuous equivalence is the Beta distribution $\mathrm{Beta}(p,k+1,N-k+1)$. In theory, p' should be drawn from $\mathrm{Beta}(p,k+1,N-k+1)$. Under a specific confidence level, say, 90%, the uncertainty width of p' is given by

$$\Delta p' = \mathrm{Beta}(0.95,k+1,N-k+1) - \mathrm{Beta}(0.05,k+1,N-k+1), \tag{2.40}$$

and $\Delta p'$ is the uncertainty width of the cumulative distribution line in the P-P plot.

[2]With 95% confidence, so it is not a hard rule.

2.2.3 Fisher matrix

Obtaining the full posterior can be time-consuming. In some cases, the focus is solely on understanding how well GW observations can constrain certain parameters, necessitating only the error bars from the posterior distribution. The Fisher matrix formalism provides a straightforward approach for this purpose [111]. Consider a maximum likelihood estimator, which is equivalent to the Bayesian posterior with a flat prior, and it gives an estimate $\bar{\boldsymbol{\theta}}$, where the derivatives of the log likelihood vanish:

$$\partial_i \ln p(\boldsymbol{\theta}|d) \mid_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} = (\partial_i h|d-h) \mid_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} = 0, \tag{2.41}$$

where ∂_i denotes the derivative with respect to the ith parameter. $\bar{\boldsymbol{\theta}}$ may differ from the true parameter $\boldsymbol{\theta}$ by an error $\Delta\theta^i$ for the ith parameter. Assume $\Delta\theta^i$ is sufficiently small, Eq. 2.41 can be written as

$$\begin{aligned} 0 &= (\partial_i h(\bar{\boldsymbol{\theta}})|d-h(\bar{\boldsymbol{\theta}})) = (\partial_i h(\bar{\boldsymbol{\theta}})|d-h(\boldsymbol{\theta})+h(\boldsymbol{\theta})-h(\bar{\boldsymbol{\theta}})) \\ &= (\partial_i h(\bar{\boldsymbol{\theta}})|n-\partial_j \Delta\theta^j h(\bar{\boldsymbol{\theta}})) \\ &= (\partial_i h(\bar{\boldsymbol{\theta}})|n) - (\partial_i h(\bar{\boldsymbol{\theta}})|\partial_j h(\bar{\boldsymbol{\theta}}))\Delta\theta^j. \end{aligned} \tag{2.42}$$

We have used $d = h + n$ and expaned $h(\boldsymbol{\theta}) - h(\bar{\boldsymbol{\theta}})$ to the linear order. Define the Fisher matrix[3]

$$F_{ij} = (\partial_i h|\partial_j h), \tag{2.43}$$

then we have

$$\Delta\theta^j = (F^{-1})^{ij}(\partial_i h(\bar{\boldsymbol{\theta}})|n). \tag{2.44}$$

Eq. 2.44 quantifies the statistical error of the maximum likelihood estimator, and reveals that the statistical error is caused by random noise. $\Delta\theta^j$ follows a multivariate Gaussian distribution since it is a superposition of Gaussian noises. Using the same trick in Eq. 2.5, we can verify that the expectation over noise

$$\langle \Delta\theta^j \rangle = 0, \tag{2.45}$$

which means the maximum likelihood estimator is *unbiased* and

$$\langle \Delta\theta^i \Delta\theta^j \rangle = (F^{-1})^{ij}, \tag{2.46}$$

[3]This definition aligns with the definition of the Fisher information matrix in statistics, which is the variance of the score, where the score is the derivative of the log likelihood with respect to parameters.

which gives the *covariance matrix* of the statistical error. Eq. 2.46 can be used to predict the statistical uncertainty of the measurement of a parameter:

$$\langle(\Delta\theta^i)^2\rangle = (F^{-1})^{ii}. \tag{2.47}$$

If there are multiple detectors, their likelihoods should be multiplied, which results in a summation of their Fisher matrices.

We should note that the Fisher matrix formalism relies on Gaussian and high-SNR approximation, as we expanded $h(\boldsymbol{\theta})-h(\bar{\boldsymbol{\theta}})$ only to the first order. Fisher matrix formalism is not valid if SNR is low or the posterior is not Gaussian (e.g. in the case of bimodal distributions), although it might have already been abused by the community [325]. Some efforts have been made to extend the Fisher matrix to higher orders to model non-Gaussian posterior, see e.g. [326]. Despite the limitations, the Fisher matrix's prediction is proved to be a lower bound of the statistical uncertainty, known as the Cramér-Rao bound [327].

2.3 Machine learning methods

We have discussed statistical methods for GW detection and parameter estimation, which rely on intensive matched filtering or waveform evaluation and can be computationally expensive. Machine learning methods in GW data analysis have received considerable interest in recent years. In this section, I will briefly introduce the basic idea of machine learning and its applications in GW astronomy.

2.3.1 Introduction to machine learning

Machine Learning (ML) is a branch of Artificial Intelligence (AI) that focuses on algorithms and statistical models that enable computers to perform specific tasks without explicit instructions. Machine learning systems are trained to recognize patterns in data and make decisions or generate outputs based on that data. This ability to learn and improve from experience is what makes machine learning a powerful tool in various domains, including image and audio recognition and generation [328, 329, 330], natural language processing [331, 332] etc, and for our purpose, GW data analysis.

To build ML model to solve a specific problem, we usually need to collect a large amount of data from the problem with which the machine uses to learn. This data can be in the form of text, images, numbers, or any other type of information pertinent to the problem at hand. For example, if the goal is to predict stock prices, the data should at least contain the historical stock price time series data. Additionally, auxiliary data such as company financial statistics or industry trends might also be collected to provide context and enable the machine to learn the underlying relationships within the data.

Moreover, it is often required to clean and transform the data into a format suitable for analysis so that the machine can learn better. This might involve removing noise, handling missing values, and normalizing the data. Data collection and pre-processing are crucial to ML models, as they have direct impacts on the applicability and performance of the model.

We should also identify which type of task we want the model to do. Common tasks include

- Regression: predict continuous numerical values (labels).
- Classification: predict categorical labels, where labels are specified in the training data.
- Clustering: without labels in the training data, group similar data points into clusters.
- Dimensionality Reduction: reduce the size of the data.
- Anomaly Detection: identify unusual patterns that do not conform to expected behavior.
- Generation: generate outputs that resemble the training dataset.

We choose the appropriate type of model for the task. Common types of ML include supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Supervised learning is the most common paradigm in ML, where the model is trained on labeled data - each training example is paired with an output label. In the stock price example, the label is the price. The goal is to learn a mapping from inputs to outputs that can be used to predict the labels of new, unseen data. Unsupervised learning deals with data that has no labels. The goal is to model the underlying structure or distribution in the data for further application. This can involve clustering data into groups with similar characteristics or reducing the dimensionality of the data. Semi-supervised learning falls between supervised and unsupervised learning. Trained on both labeled and unlabeled data, it aims to combine the advantages of both approaches. Reinforcement learning uses an agent to learn and make decisions by performing actions in an environment to maximize cumulative reward. Unlike supervised learning, where the model learns from a fixed dataset, reinforcement learning learns from the consequences of actions through trials and errors.

We can then assemble the model. In this thesis, I will focus on Neural Networks (NNs), a specific subset of ML models inspired by the structure and function of the human brain. NNs consist of interconnected layers of nodes (neurons) that work together to process input data and produce output predictions. In other words, we build the

model by connecting a series of manipulations to the input data which finally leads to the output we need. The manipulations include arithmetical and matrix calculations, analytical functions, convolution and so forth. Here I give some examples of common NN structures:

- Multi-Layer Perceptron (MLP). This is the most basic type of NN, also known as the fully connected, feed-forward or dense network. It consists of an input layer, one or more hidden layers, and an output layer. Given the input as an array, each layer performs a linear transformation (which could change the dimension) and an activation function. The activation function needs to be non-linear, easy to compute and differentiable, such as Rectified Linear Unit (ReLU), `tanh`, and sigmoid function. MLP is suitable for any array-like input but is inefficient for high-dimensional datasets. It is often used as the basic building block in many complex networks.

- Convolutional Neural Network (CNN). It applies a filter (kernel) to local regions of the input, extracting the local structures (feature maps). The usual follow-ups are pooling (taking maximum or average) to reduce the dimensions of the feature maps and a MLP to reshape the output. CNN is designed for processing grid-like data such as images but is also suitable for 1D input.

- Residual Network (ResNet) [333]. It is a modification to the overall architecture of a standard neural network, such as MLPs and CNNs. In addition to processing the data with a common network, the input data is then directly added (sometimes a linear reshape is required) to the processed data. These residual connections enable the network to learn residual functions with respect to the input, which makes it easier to train very deep networks.

- Transformer [334]. Transformers are a class of deep learning models known for their ability to capture long-range dependencies in sequential data through the self-attention mechanism. Unlike traditional neural networks that process inputs sequentially or in fixed local contexts, transformers allow each element of the input to interact with every other element directly. This is achieved by computing attention scores, which determine the relevance of one part of the input to another, enabling the model to build rich contextual representations. The transformer encoder consists of stacked layers of self-attention and feed-forward neural networks. This architecture is particularly powerful for handling complex patterns and dependencies in the data.

The possibilities for building NNs are endless and it is impossible to give a comprehensive overview for all NNs. In Chapter 6, I will further introduce *normalizing flows*, a generative model that can generate posterior samples for parameter estimation.

After setting the model structure, the model needs to be trained and validated before use. The collected data should be split into at least two subsets, one for training and one for validation. The model will be trained only on the training set, and we assess the model's performance using the validation set to ensure it generalizes well to new data. We should assign a loss function to quantify the performance. For example, if the model learns to fit a continuous function, the loss can be a mean square error of the prediction; if the model learns to generate a distribution, the loss function can be the Kullback-Leibler Divergence (KLD) or cross-entropy.

The training is essentially optimizing the parameters within the network, such as the linear transformation matrices in the MLP and the kernel weights in CNN. After initiating the parameters, we take a *batch* of data which will be running in parallel, pass the data through the network and generate the outputs and the mean loss of the batch. We then calculate the gradients of the loss with respect to the network's parameters using *back propagation*, i.e., applying the chain rule of derivatives to propagate the error backward through the network. All parameters will be updated using an optimization algorithm such as SGD and Adam [335]. The optimizer adjusts the weights to minimize the loss based on the calculated gradients: $\boldsymbol{\phi}_{i+1} = \boldsymbol{\phi}_i + \eta \nabla L\left(\boldsymbol{\phi}_i\right)$[4], where $\boldsymbol{\phi}_i$ is the parameters of the network and η is usually known as the *learning rate*. We repeat this process so that the model is iteratively optimized, until the loss function stops to decrease, which indicates convergence. Training data will be reused in this process: an entire loop over all data once is referred to as an *epoch*. By following these steps, the machine learning model is trained to perform effectively on the task at hand. The effectiveness is gauged by its performance on the validation set, and once satisfactory performance is achieved, the model can be deployed for real-world applications.

2.3.2 Applications in gravitational wave astronomy

Machine learning has been applied to a wide range of problems in GW astronomy and has shown great potential compared to traditional methods [336]. In this section, I will give a brief overview of applications of ML in GW astronomy, with a specific focus on ground-based detectors.

Waveform modeling

NNs excel in modeling highly complex functions, which makes them good candidates when a surrogate function is needed in GW waveform modeling. Additionally, leveraging GPU acceleration enhances the efficiency of waveform generation, contrasting with traditional CPU-based platforms. Ref. [337] exemplified the application of NN in constructing

[4]This is the simplest form of gradient descent. It could take more complex forms inside optimizers.

reduced-order models of GW waveforms, achieving decent accuracy and speed. Extending this capability to precessing systems, Ref. [265] demonstrated simple MLPs perform exceptionally well in the high dimensional parameter space where traditional spline methods could struggle. Ref. [275] made use of the auto-differentiation algorithm `JAX` [338] and enabled the direct calculation of the derivatives of the waveform, which could bring significant speed-up in data analysis [339]. ML methods also have the potential to improve the robustness of waveforms: Refs. [266, 267, 268, 269] used Gaussian progress regression to build BBH models that provide waveform uncertainty while evaluating the waveform. Beyond CBC source, Generative Adversarial Networks (GAN) has been applied to model burst signals [340].

Detector noise

GW data often includes numerous glitches that can significantly impact data analysis. The identification and removal of these glitches are critical in ensuring the reliability of GW data analysis, and classifying glitches could bring a better understanding of the noise sources. `GravitySpy` [118] uses NNs to identify and classify glitches directly from GW data. `iDQ` [341], on the other hand, detects noise artifacts in GW detectors based on auxiliary channels that are insensitive to GWs. Additionally, GAN is used to simulate transient noise bursts [342].

Detection

Instead of using matched filtering to detect GWs, considerable efforts have been directed towards ML-based detection of GWs, including CBC signals [343, 344, 345, 346, 347], burst signals [348, 349, 350], continuous waves [351, 352], and stochastic background [46]. It is reported that ML-based detection is faster than statistical methods without much loss of sensitivity. Notably, the `MLy` [346] pipeline, an ML-based unmodelled search pipeline participated in O4 detection of LVK.

GW detections can be further processed by ML algorithms. For example, NNs are able to reconstruct the GW signal from the data (known as denoising) [347, 353, 354] and achieved comparable performance as wavelet reconstruction. Ref. [355] developed a real-time framework to distinguish between astrophysical events and instrumental artifacts using public real-time detection data, which is helpful to the decisions on EM follow-ups.

Parameter estimation and astrophysical inference

Parameter estimation is the most computationally expensive part of CBC data analysis. ML-based full parameter estimation of the source parameters has been achieved with generative models. Ref. [356] took time domain GW data and used GAN to generate posterior

samples, and showed that ML models can perform correct Bayesian analysis in CBC parameter estimation. However, GAN is not optimized with the exact likelihood, which poses challenges for achieving optimal convergence. Ref. [357] proposed `DINGO` based on normalizing flow instead of GAN, which can be trained on the exact likelihood. `DINGO` takes frequency domain data alongside the noise so it can adapt to the noise variations in the real detector. With importance sampling [358], `DINGO` achieved negligible difference with full Bayesian parameter estimation on real GW events. Recently, `DINGO` framework has been extended to eccentric sources [142], enabling full parameter estimation with computationally expensive EOB eccentric waveforms, which is prohibitively slow for traditional methods. A similar network structure has been applied to overlapping signals in the 3G detectors and performed 30-dimension parameter estimation [359].

Other than the end-to-end parameter estimation, NNs could help accelerate the stochastic samplers used in parameter estimation. `Nessai` [274] is a nested sampling algorithm that uses normalizing flow to sample the proposal distribution and achieves faster performance compared to traditional methods. `FlowMC` [339] also makes use of normalizing flows, enabling full Bayesian inference within minutes combined with differentiable waveforms and relative binning.

NNs are also used in astrophysical inference. Ref. [360] proposed a localization algorithm that takes the matched filtering SNR timeseries as input and is able to localize the GW source in real-time detection. Ref. [361] demonstrated the ability to constrain the equation of state of neutron stars using normalizing flows within one second. On the population level, NN could speed up the inference for astrophysical population models and cosmological models [362, 363, 364]. Moreover, non-parametric methods are used to explore the population properties beyond parametrized astrophysical models [365, 366, 367].

2.4 Challenges in data analysis

With the improvement of detector sensitivity, we are entering the era of high-precision GW astronomy. This means more events will be detected and the high-SNR "golden" events will be more frequent. While this is exciting from a physics perspective, these detections will bring challenges to data analysis, potentially complicating the process and leading to erroneous conclusions. In this section, I will discuss the main challenges in parameter estimation of CBC signals in ground-based GW detectors: systematic errors and computational cost.

2.4.1 Systematic errors

We construct the likelihood for GW parameter inference with the Gaussian noise assumption, i.e., we assume the data is Gaussian noise after removing the waveform. This leads

to the statistical error Eq. 2.44, which states that the maximum likelihood point could deviate from the true value due to the noise but the estimator is unbiased on the ensemble level. However, it is not possible to perfectly remove the signal from data. The waveform models used in data analysis are approximations of the true solution and could miss some physics such as the orbital eccentricity, resulting in residual strains after signal subtraction. In fact, the residual strains are not limited to waveform residuals. Glitches can also contribute, though they can often be identified and removed. For the 3G detectors, it is expected to detect multiple signals in a narrow time window, leading to overlapping signals that may leave residual strains impacting each other.

The residual strains could cause *systematic* errors in parameter estimation. The generic formalism for estimating systematic errors in PE was first proposed in [368] and widely used (e.g. [369, 370]), and was generalized and validated against full PE by [371]. Mathematically, the maximum likelihood estimate $\bar{\boldsymbol{\theta}}$ differs from the real parameter $\boldsymbol{\theta}_{\rm real}$, and the waveform model $h_{\rm m}$ differs from the true waveform h. We have Eq. 2.41

$$\partial_i \ln p(\boldsymbol{\theta}|d) \mid_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} = (\partial_i h|d-h) \mid_{\boldsymbol{\theta}=\bar{\boldsymbol{\theta}}} = 0. \tag{2.48}$$

The data d is known, but real parameter $\boldsymbol{\theta}_{\rm real}$ and the GW signal in the detector $h(\boldsymbol{\theta}_{\rm real})$ is unknown. In practice, they are replaced by a waveform model $h_{\rm m}(\bar{\boldsymbol{\theta}})$. By doing this, errors are introduced to $d-h$:

$$\begin{aligned} & d - h_{\rm m}(\bar{\boldsymbol{\theta}}) \\ &= d - h(\boldsymbol{\theta}_{\rm real}) + h(\boldsymbol{\theta}_{\rm real}) - h_{\rm m}(\boldsymbol{\theta}_{\rm real}) + h_{\rm m}(\boldsymbol{\theta}_{\rm real}) - h_{\rm m}(\bar{\boldsymbol{\theta}}) \\ &= n + \delta H + \Delta\theta^j \partial_j h_{\rm m}. \end{aligned}$$

The first term n is what $d-h$ is supposed to be: the noise in the detector. The second term $\delta H = h(\boldsymbol{\theta}_{\rm real}) - h_{\rm m}(\boldsymbol{\theta}_{\rm real})$ is the excess strain which represents the difference between real signal(s) in the data and the model used to subtract signals. The third term comes from the imperfect measurement of signal parameters due to statistical noise, and is given by the linear expansion of $h_{\rm m}(\boldsymbol{\theta}_{\rm real}) - h_{\rm m}(\bar{\boldsymbol{\theta}})$, where $\Delta\theta^j$ is the error of the j'th parameter from the maximum likelihood estimator. Substituting Eq. 2.49 into Eq. 2.48 and approximate all derivatives at $\bar{\boldsymbol{\theta}}$, and we get

$$\Delta\theta^i \approx (F^{-1})^{ij}(\partial_j h_{\rm m}|n+\delta H) = \Delta\theta^i_{\rm stat} + \Delta\theta^i_{\rm sys}, \tag{2.49}$$

where $\Delta\theta^i_{\rm stat}$ is the statistical error Eq. 2.44, and

$$\Delta\theta^i_{\rm sys} = (F^{-1})^{ij}(\partial_j h_{\rm m}|\delta H) \tag{2.50}$$

is the systematic error for a maximum likelihood estimator because $\langle \Delta\theta^i_{\rm sys} \rangle \neq 0$. Any effect

that contributes to δH could be a source of systematic bias in parameter estimation.

The SNR ρ is proportional to waveform amplitude h, and we have $(F^{-1})^{ij} \propto 1/\rho^2$. Therefore,

$$\Delta\theta^i_{\rm stat} \propto \frac{1}{\rho}, \tag{2.51}$$

which means the statistical uncertainties shrink with the increase of SNR, in other words, parameters can be constrained better in high-SNR events. However, δH may also be proportional to the waveform amplitude h (for instance, the waveform residuals), which leads to

$$\Delta\theta^i_{\rm sys} \propto 1, \tag{2.52}$$

i.e., the systematic is a constant! This means that while high-SNR improves the *precision* of the measurement, it is the residual strain that ensures the *accuracy*. Systematic errors could dominate the parameter estimation in high-SNR scenarios if the residual strain is not well controlled, leading to a precise but inaccurate parameter estimation. This could lead to some incorrect scientific conclusions as it rules out the true value. It is therefore necessary to develop analysis methods that are robust against systematic errors, or carefully assess the potential systematic errors in the analysis, especially in the high-SNR scenarios.

2.4.2 Computational cost

The large number of events and high SNRs present challenges not only in terms of robustness but also in terms of computational burden. It is estimated that 3G detectors are able to detect $> 10^5$ CBC events per year [171]. With current methods used in analyzing LVK catalogs, parameter estimation of these events could take tens of millions of CPU hours. This estimate might be conservative because many of the events are long BNS signals and high SNR signals that are particularly slow to analyze – some even prohibitively so for current sampling methods. While machine learning methods seem promising in relieving the computational burden, substantial work remains to adapt these algorithms for real-world applications.

2.5 Summary and the goal of this thesis

In this chapter, I reviewed the basic data analysis methods of CBC sources for the ground-based detectors. On the detection side, I explained how matched filtering is constructed and how complex SNR timeseries is calculated, and gave a brief overview of detection pipelines. For the parameter estimation, I introduced the standard Bayesian inference framework and Fisher matrix framework. I also briefly introduced machine learning algorithms and their applications in GW astronomy. In the end, I mentioned two main

challenges in data analysis: systematic errors and computational burden.

The goal of this thesis is to address these two challenges for current and future GW observations. In the first half of the thesis (Chapter 3 and 4), I will evaluate the seriousness of systematic errors in current and future observations. Using testing GR as an example, I will first show the consequences of systematic errors, including waveform systematics and overlapping signals for the 3G detectors in Chapter 3. Next, in Chapter 4, I will propose a method to assess GW waveform accuracy and apply it to LVK open data to evaluate the systematics that already exist in current data analysis, and predict the accuracy improvements required for the 3G detectors.

In the second half (Chapter 5 and 6), I will develop efficient data analysis algorithms to alleviate the computation burden in the 3G era. More specifically, I will present a semi-analytical localization method in Chapter 5, and a ML-based parameter estimation method in Chapter 6, to analyze long BNS signals in the 3G detectors. The former aims to provide prompt sky directions in real-time and early-warning detection for capturing EM counterparts, while the latter aims to generate full parameter estimation for long BNS signals on a catalog level with minimal computational cost.

Chapter 3

Accumulating errors in tests of general relativity

Observations of Gravitational Waves (GWs) from Compact Binary Coalescences (CBCs) provide powerful tests of General Relativity (GR), but systematic errors in data analysis could lead to incorrect scientific conclusions. This issue is especially serious in the third-generation GW detectors in which the Signal-to-Noise Ratio (SNR) is high and the number of detections is large. In this chapter, we investigate the impacts of overlapping signals and inaccurate waveform models on tests of GR. We simulate mock catalogs for Einstein Telescope and Cosmic Explorer and perform parametric tests of GR using waveform models with different levels of inaccuracy. We find the systematic error in non-GR parameter estimates could accumulate toward a false deviation from GR when combining results from multiple events, although a Bayesian model selection analysis may not favour a deviation. Waveform inaccuracies contribute most to systematic errors, but multiple overlapping signals could magnify the effects of systematics due to the incorrect removal of signals. We also point out that testing GR using selected "golden binaries" with high SNR is even more vulnerable to false deviations from GR. The problem of error accumulation is universal; we emphasize that it must be addressed to fully exploit the data from third-generation GW detectors, and that further investigations, particularly in waveform accuracy, will be essential.

This chapter is organized as follows. We will introduce the systematic errors in testing GR in Sec. 3.1, including the Parametrized Post-Newtonian (PPN) formalism, the method of predicting systematic errors, and an example event. In Sec. 3.2 we will simulate mock catalogs for the 3G detectors and investigate how errors accumulate when combining multiple events. A summary and discussions will be given in Sec. 3.3.

3.1 Systematic errors in testing GR

3.1.1 Status and prospects of testing GR with GWs

The observation of GWs from CBCs provides an ideal means of testing of GR in the strong-field regime [25, 155, 156, 157, 158, 372, 373]. The latest GW event catalogs contain nearly 100 CBC events [127, 128], based on which various tests of GR have been performed [25, 158]. No concrete evidence of a deviation from GR has been found yet, but unprecedented constraints have been placed on possible violations of the theory. In the coming decades, the third-generation (3G) ground-based GW detectors (i.e. the Einstein Telescope [166]) and Cosmic Explorer [168] are expected to detect $\mathcal{O}(10^5)$ CBC events per year, with SNR up to thousands [374, 375, 376, 377, 378]. Since the statistical uncertainty of parameter estimates shrinks when the SNR increases, and when a catalog of events are combined, observations from 3G GW detectors are expected to be able to obtain much tighter constraints on gravity theories.

However, this inspiring prospect of an enlarged detection catalog and higher SNRs brings with it many difficulties in data analysis. For the purpose of testing GR (and any other theories), one needs to ensure that the systematic errors are small, so that the analysis will not favor the wrong theory and cause a false alarm (or false dismissal). Parameterized tests of GR [379] suffer from the same problems as Parameter Estimation (PE) in general, which has been investigated in many works, (e.g. [368, 371]. For instance, inaccurate waveform models may have already caused some tensions in current GW observations [2, 380] and are expected to be more important in future high SNR detections [368, 381, 382]. Additionally, the 3G detectors with their improved low-frequency sensitivity are able to observe multiple signals at the same time. Detected overlapping signals can not be perfectly removed from the data due to uncertainties in parameter estimation, and could have non-negligible impacts on PE when the merger times of overlapping signals are close [371, 375, 376, 377, 383, 384]. The undetected overlapping signals, i.e., the signals that are too faint to be detected may also contribute to the systematic error [371, 385]. These errors are inevitable in 3G detectors, and repeated biased estimations for each event might lead to a wrong conclusion in the catalog-level analysis [386, 387].

The aforementioned works mainly focus on case studies for single events, or include only one type of systematic error. In this chapter, we aim to perform a more comprehensive investigation of systematic errors at the catalog level, including interactions between different types of systematics, namely, waveform inaccuracies and overlapping signals.

3.1.2 Estimating systematic errors

We have introduced the method of estimating systematic errors in PE in Sec. 2.4, and argued that they are caused by the residual strain δH via Eq. 2.50. Inaccurate waveforms and overlapping signals can both contribute to this term. The excess strain from inaccurate waveforms can be simply written as

$$\delta H_{\rm WF} = h_{\rm m}(\boldsymbol{\theta}_{\rm real}) - h(\boldsymbol{\theta}_{\rm real}) \approx h_{\rm m}(\bar{\boldsymbol{\theta}}) - h(\bar{\boldsymbol{\theta}}) \tag{3.1}$$

where $\boldsymbol{\theta}_{\rm real}$ is the real source parameter and $\bar{\boldsymbol{\theta}}$ is the maximum likelihood parameter. The approximation comes from the fact that the error would be a higher-order term. The waveform systematic is given by

$$\Delta\theta^i_{\rm WF} = (F^{-1})^{ij}(\partial_j h_{\rm m}|\delta H_{\rm WF}), \tag{3.2}$$

where $(F^{-1})^{ij}$ is the inverse of the Fisher matrix Eq. 2.46.

The estimates for overlapping signal residual strains are more complex. Overlapping signals can be classified into two types: detected signals and undetected signals (confusion signals). The former is strong enough to be detected and should be subtracted from data in the analysis for other signals (or, the "main" signal)[1]. The latter, however, is too faint to be recognized by the detection pipeline and may have an unnoticed impact on PE. In this work, the network SNR threshold for detection is set to 8, under which GWs are assumed to be undetected. Moreover, it is known that the correlation between signals is not strong unless the merger times are very close (typically < 1s); in this work, we regard two signals as "overlapping" only if the merger time difference $|\Delta t| < 4$s, which captures the most influential neighbours of a signal.

If a signal is detected, it will still contribute to excess strain since we cannot perfectly remove it from the data. The excess strain after imperfect removal is

$$\delta H_{\rm DO} = h'(\boldsymbol{\theta}'_{\rm real}) - h'_{\rm m}(\bar{\boldsymbol{\theta}}) \approx \Delta\theta'^i \partial_i h'_m + \delta H'_{\rm WF}, \tag{3.3}$$

where $'$ denotes variables of the detected overlapping signal. The first term arises from the inaccurate estimation of parameters for the overlapping signal, which is random since the error is partly caused by the random noise, although other factors, such as waveform inaccuracies and overlapping signals also contribute to it. As a conservative estimation and following [371], we ignore waveform systematic errors in $\Delta\theta'^i$ (i.e., assuming $\Delta\theta'^i$ is merely caused by noise, which tends to underestimate it), and adopt the lowest order approximation for its correlation with the main signal. Substituting it into Eq. 2.49, one

[1] It is also possible to do a joint parameter estimation for all existing signals, see [384].

obtains the covariance of the first term in the systematic error Eq. 3.3

$$< \Delta\theta^i_{\rm DO1}\Delta\theta^j_{\rm DO1} >= \left(\boldsymbol{F}^{-1}\boldsymbol{F}_{\rm mix}\boldsymbol{F}'^{-1}(\boldsymbol{F}_{\rm mix})^{\rm T}(\boldsymbol{F}^{-1})^{\rm T}\right)^{ij}, \tag{3.4}$$

where $(F_{\rm mix})_{ij} = (\partial_i h|\partial_j h')$ encodes the correlation between two signals and $F'_{ij} = (\partial_i h'|\partial_j h')$ is the Fisher matrix of the overlapping signal. The second term in Eq. 3.3 represents the inaccurate waveform model we use to subtract signals, and can be calculated the same way as the waveform systematic, yielding

$$\Delta\theta^i_{\rm DO2} = (F^{-1})^{ij}(\partial_j h_{\rm m}|\delta H'_{\rm WF}) \tag{3.5}$$

In this work, the systematic error from detected overlapping signals is calculated as $\Delta\theta^i_{\rm DO2}$ plus a random sample drawn from a multivariate Gaussian distribution with covariance matrix Eq. 3.4 and zero mean, i.e.,

$$\Delta\boldsymbol{\theta}_{\rm DO1} \sim N(\mathbf{0}, \boldsymbol{F}^{-1}\boldsymbol{F}_{\rm mix}\boldsymbol{F}'^{-1}(\boldsymbol{F}_{\rm mix})^{\rm T}(\boldsymbol{F}^{-1})^{\rm T}), \tag{3.6}$$

and

$$\Delta\boldsymbol{\theta}_{\rm DO} = \Delta\boldsymbol{\theta}_{\rm DO1} + \Delta\boldsymbol{\theta}_{\rm DO2}. \tag{3.7}$$

We do not simply take the expectation for $\Delta\boldsymbol{\theta}_{\rm DO1}$ because it may have a different sign with other systematic errors and cancel with them. For more than one detected overlapping signals, Eq. 3.3 can be extended by defining h' as the summation of all GWs in the data [371], which enlarges the dimension of $\boldsymbol{F}_{\rm mix}$ and $\boldsymbol{F}'$.

The undetected overlapping signal simply contributes to systematic error by

$$\delta H_{\rm UO} = \sum_{\rm undetected} h''(\boldsymbol{\theta}''_{\rm real}), \tag{3.8}$$

$$\Delta\theta^i_{\rm UO} = (F^{-1})^{ij}(\partial_j h_{\rm m}|\delta H_{\rm UO}), \tag{3.9}$$

which is accessible in our simulation but unknown in real data analysis. Here $''$ denotes the variables of the undetected overlapping signal.

3.1.3 Parameterized post-Newtonian formalism

Testing GR

The test of parameterized post-Newtonian coefficients is a generic formalism for finding deviations from GR, initially proposed by [388] and further developed for application with Bayesian inference [389], and later applied to catalogs of real GW observations, most recently in [25]. We use the waveform model `IMRPhenomPv2` [237, 251], whose phase is char-

acterized by a set of parameters $\{p_i\}$, including inspiral phase parameters $\{\phi_0, \ldots, \phi_7\}$ and $\{\phi_{5l}, \phi_{6l}\}$, phenomenological coefficients $\{\beta_0, \ldots, \beta_3\}$, and merger-ringdown parameters $\{\alpha_0, \ldots, \alpha_5\}$. Deviations $p_i \to (1 + \delta\hat{p}_i)p_i$ are introduced as the violations of GR; $\delta\hat{p}_i = 0$ reproduces GR. In this framework, testing GR is reduced to estimating the testing parameters $\delta\hat{p}_i$. Although a specific modified gravity theory could bring deviations in more than one testing parameter, previous works have shown that including one testing parameter at once is enough to detect violations. In fact, it can be more efficient to find violations from GR this way because it avoids the correlations between testing parameters and GR parameters [379, 390]. In this work, we choose $\delta\hat{\phi}_0$, the perturbation on the 0PN phase, as the example testing parameter. We assume GR is the correct theory and focus on whether the PPN test falsely indicates deviations of GR.

We restrict our Fisher matrix analysis to a subset of the full signal parameters, to avoid computational issues. Parametrized deviations of the type we consider have a direct effect on the phasing of the signal, so in addition to $\delta\hat{\phi}$ we must include the other parameters that do the same: chirp mass $\mathcal{M}$ and mass ratio q, as well as the time of coalescence t_c. The full 6-dimensional space of spin configurations is known to bring ill-conditioned Fisher matrices [171] due to correlations between parameters, and because of the prior bounds on angular parameters, results can be misleading even when they can be computed. We therefore use only the effective spin $\chi_{\rm eff}$ to capture the dominant effect of (aligned) spin on the waveforms. We include this by forcing the two aligned spin components to contribute equally to $\chi_{\rm eff}$, which allows us to treat it as a single parameter. We neglect to include extrinsic parameters in the Fisher matrix, effectively assuming they are measured precisely. Since these do not have a frequency-dependent effect on the phase, we do not expect them to be highly correlated with the intrinsic parameters. Our choice captures the parameters that appear in the leading PN term and the corresponding PPN modifications, as well as the decisive parameter in the analysis of overlapping signals, t_c. Other parameters are randomly generated but are treated as perfectly known. Setting parameters to their injection values excludes their contributions to both statistical and systematic errors in PE. For instance, if we removed the effective spin from our calculation, we would obtain tighter statistical and systematic errors, because its correlation with other parameters is removed in that case [391]. Considering realistic PE in the future in which all parameters are included, the correlation between parameters may make posteriors wider and systematic bias larger. However, due to the linear expression in Eq. 2.49, we expect the two changes to be proportional and our conclusion will not change significantly under this simplification.

Mimicking waveform inaccuracies

The PPN formalism is also used to mimic waveform systematics. We induce a non-zero $\delta\hat{\beta}_2$ to represent inaccurate waveform models based on the following considerations. To

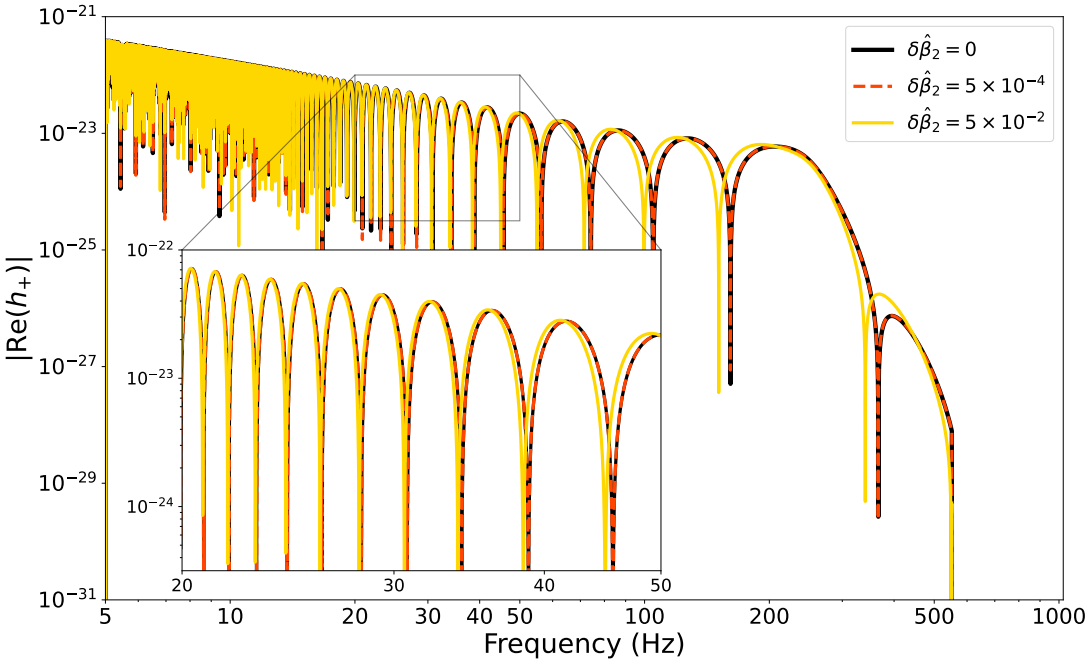


Figure 3.1: The absolute value of real part of plus polarization from a non-spinning BBH with $\mathcal{M}_c = 30.69\ \mathrm{M}_\odot$, $q = 0.88$, in the frequency domain. Waveforms with different $\delta\hat{\beta}_2$ are shown in different colors and linestyles. The intermediate region of this system starts around 50 Hz, which is consistent with where waveform difference appears in the plot.

reduce potential correlations with $\delta\hat{\phi}_0$, we exclude testing parameters for the inspiral stage. Correlation between the testing parameter and the waveform systematic parameter may undermine the generality of the illustration. To make sure the testing parameter has enough influence on the waveform, we do not choose parameters for the merger-ringdown stage which only includes the last few cycles. Therefore, we look for parameters in the intermediate region which is described by $\delta\hat{\beta}_i$ [251]. $\delta\hat{\beta}_0$ and $\delta\hat{\beta}_1$ bring global phase shift and time shift in this region respectively, so $\delta\hat{\beta}_2$ is the dominant testing parameter that encodes physical (frequency-dependent) modifications.

We assume $\delta\hat{\beta}_2 = 0$ is our model waveform, while the "real" waveform could have $\delta\hat{\beta}_2 = 0$, 5×10^{-2}, or 5×10^{-4}. The first case means our model waveform is perfect, and all systematic errors will come from overlapping signals. The second case generates waveform mismatches around $10^{-4}-10^{-3}$, which corresponds to the current waveform accuracy [392, 393]. The last case produces mismatches around $10^{-7} - 10^{-6}$ and corresponds to the expectations for future waveform accuracy [2, 382]. A comparison of the three types of waveforms is shown in Fig. 3.1. We show an example of a non-spinning BBH merger with $\mathcal{M}_c = 30.69\ \mathrm{M}_\odot$ (in the detector frame), and $q = 0.88$ whose intermediate region starts around 50Hz. The mismatches are 3×10^{-7} and 2×10^{-3} between $\delta\hat{\beta}_2 = 0$ and $\delta\hat{\beta}_2 = 5\times10^{-4}, 5\times10^{-2}$, respectively.

False deviations: a case study

We will use $\sigma_i = \sqrt{(F^{-1})^{ii}}$ as the statistical uncertainty. The systematic error is given by the combinations of Eqs. 3.2, 3.7, and 3.9. We define the error ratio between systematic and statistical errors as

$$\mathcal{R}(\theta_i) = |\Delta\theta_{\text{i, sys}}/\sigma_i|. \tag{3.10}$$

We consider that the PPN test coefficient is subject to false deviations from GR when $\mathcal{R}(\delta\hat{\phi}_0) > 1$, i.e., the systematic error dominates the parameter estimation.

Here we present an example event to investigate the effect of a detected or undetected overlapping signal. The main signal is from a BBH with $\mathcal{M}_{\rm c} = 32\mathrm{M}_\odot$ (in the detector frame), $q = 0.9$, $\chi_{\rm eff} = 0.2$ and network SNR of 27. The overlapping signal is an equal mass BBH with $\mathcal{M}_{\rm c} = 20\mathrm{M}_\odot$ and $\chi_{\rm eff} = 0.1$. We scale its SNR from ~ 26 down to $\lesssim 8$ to make it detectable or undetectable. We vary the merger time difference (by 0.01 s per step) and calculate the total systematic error with different waveform models. Note that, throughout this section, the "systematic error" refers to that of the testing parameter $\delta\hat{\phi}_0$, and is denoted as $\Delta\theta_{\rm sys}$.

The error ratio for this example event is shown in Fig. 3.2, including an illustration of the waveforms. The error from the overlapping signal oscillates when Δt changes due to the repeating alignments and misalignments of phases of the two GWs. The overlap error is not symmetric around $\Delta t = 0$ because the two waveforms are not symmetric, but the peak is always located in the region $|\Delta t| \leq 1$s, meaning the overlapping signal only produces a large influence when two mergers are very close. Waveforms in the last row show how the main signal is modulated by overlapping signals. Around $\Delta t \sim 0$, the confusion signal has a larger impact than waveform systematics, so it dominates the systematic error. The detected signal changes the signal significantly, but it is then subtracted from the data and therefore produces fewer residual strains. When $|\Delta t|$ is large, it is waveform inaccuracy that dominates the systematic error. These characteristics are consistent with previous works [371, 375, 376, 377].

It is possible for undetected signals to produce significant systematic errors in our simulation, especially when the merger time is close. However, when the waveform is not accurate, the detected overlapping signal produces a more stable systematic error regardless of the merger time difference because the waveform systematic is always involved in signal subtraction. This will be discussed in more detail on the catalog level in the next section.

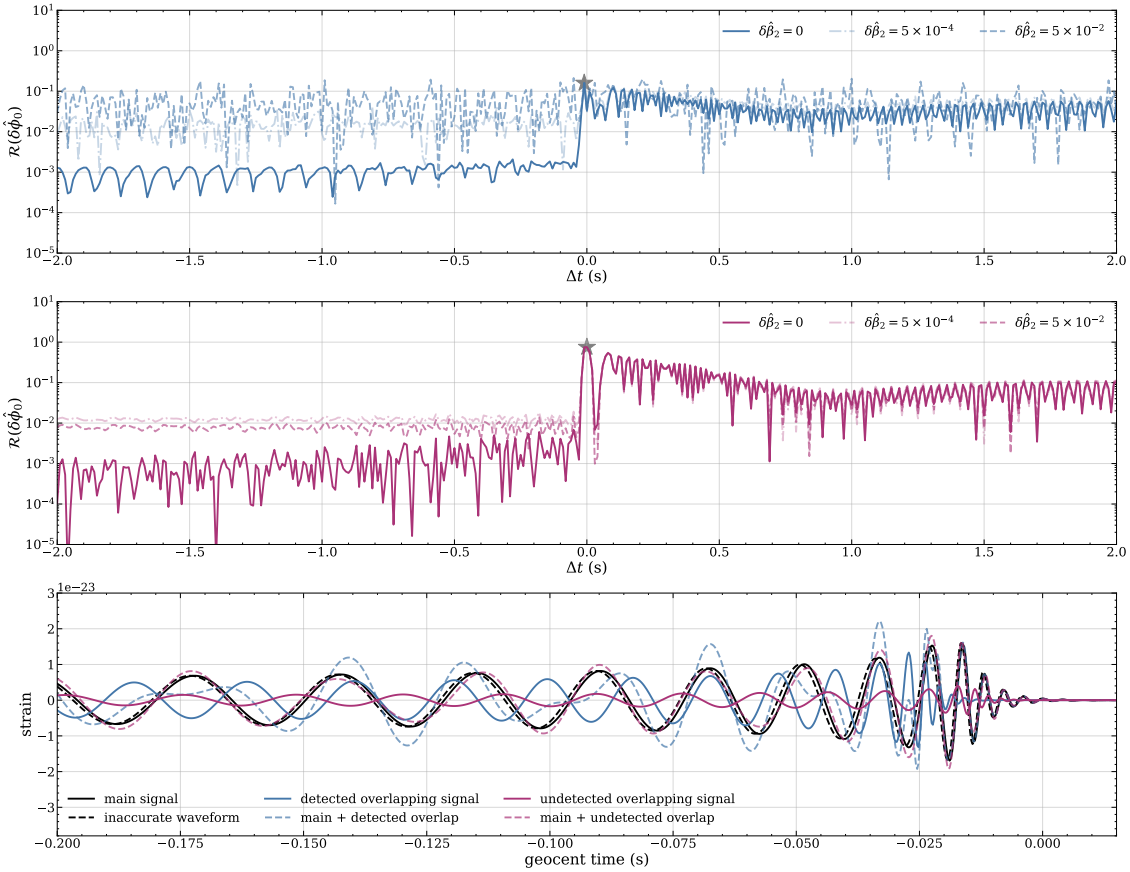


Figure 3.2: Uppermost and middle rows: The error ratio of $\delta\hat{\phi}_0$ varies with merger time difference. The main signal has (detector frame) $\mathcal{M}_c = 32\,M_\odot$, $q = 0.9$, $\chi_{\rm eff} = 0.2$ and SNR of 27. The overlapping signal is an equal mass BBH with $\mathcal{M}_c = 20\,M_\odot$ and $\chi_{\rm eff} = 0.1$. The SNR of the overlapping signal is adjusted by changing its luminosity distance: the detected overlap is shown in the upper panel, and the undetected in the lower one. We use three kinds of waveforms explained in Sec. 3.1.3: perfect waveform (solid line), "current" waveform (dashed line), and "future waveform" (faint dotted-dashed line).
Bottom row: waveforms of the main and overlapping signals and their superposition. Merger times of overlapping signals are chosen to maximize their influences, as marked by grey stars in the first two rows. Inaccurate waveform in the $\delta\hat{\beta}_2 = 5\times10^{-2}$ case is also plotted for comparison.

3.2 Catalog simulations and results

3.2.1 Mock catalogs

We are interested in error accumulation effects on the catalog level, so we need a mock catalog of CBC events. We consider BBH and BNS sources, and assume their rate distribution in redshift z follows the analytical approximation [378]

$$R_{\rm GW}(z) = \frac{a_1 e^{a_2 z}}{e^{a_3 z} + a_4} {\rm Gpc}^{-3}{\rm yr}^{-1}, \tag{3.11}$$

which is then converted to observable event rate by multiplying a factor $\frac{1}{1+z}\frac{dV_c}{dz}$. Here V_c is the comoving volume and we employ Planck15 cosmology [394]. Note that "observable" GWs need to achieve an network SNR of 8 to be "detectable". $a_{\{1,2,3,4\}}$ are model parameters. We set $a_2 = 1.6, a_3 = 2.1, a_4 = 30$ to mimic a peak at $z \sim 2$. a_1 is scaled based on local merger rate given by [149] ($\mathcal{R}_{\rm BNS} = 320^{+490}_{-240}$ and $\mathcal{R}_{\rm BBH} = 23.9^{+14.3}_{-8.6}{\rm Gpc}^{-3}{\rm yr}^{-1}$) such that $R_{\rm GW}(z=0) = \mathcal{R}_{\rm BNS/BBH}$. We choose three values for a_1 which corresponds to lower, median, and higher estimation of local merger rate, respectively.

The masses of BBHs are generated by the PowerLaw + Peak model in [149], while all BNS systems are set to be same: $1.45 + 1.4 M_\odot$, $\Lambda_1 = \Lambda_2 = 425$. The effective spin follows the Gaussian distribution in [149], with mean of 0.06 and standard deviation of 0.12. `IMRPhenomPv2_NRTidal` [256] is used to generate BNS waveforms with the same $\delta\hat{\beta}_2$ as BBH. We will perform tests of GR with all BBH events and use BNS events as a background: BNS events are only involved in the calculation as overlapping signals. We assume isotropically distributed inclination and source sky direction; and uniformly distributed coalescence time, phase, and polarization angle.

A summary of low, median, and high merger rates catalogs is shown in Tab. 3.1. It shows that most BBH events will not have an overlapping signal near their merger time, which implies overlapping signals contribute to systematic errors less frequently than waveform systematics. With our ET+CE configuration, the numbers of the two kinds of overlaps are close. However, if the number of detectors is less than assumed, or detector sensitivities are lower than designed, some of the detected overlaps would become undetected, and vice versa. The unnoticeable confusion background has drawn attention in recent works [385, 395] and needs further investigation. Compact binaries formed by Pop III stars (which we have ignored) could also contribute to the confusion background. However, according to the model in [378], the numbers of observable Pop III binaries of B17 and K16 models per year are roughly 40000 and 180000 respectively, which is much lower than the BNS background.

Several simplifications have been adopted in our mock catalog: we regard BNS as a background and use only BBH as the test source; we ignore neutron star-black hole

	# of observable binaries		Detected overlaps on BBH events		Undetected overlaps on BBH events	
	BBH	BNS	# of overlaps	# (fraction) of events	# of overlaps	# (fraction) of events
Low	56526	286088	0	53118 (95%)	0	54067 (96%)
			1	2847 (5.1%)	1	1936 (3.5%)
			2	74 (0.13%)	2	37 (0.066%)
			3	2 (0.0040%)	3	1 (0.0018%)
Median	88300	1144354	0	73200 (84%)	0	76270 (87%)
			1	13125 (15%)	1	10461 (12%)
			2	1093 (1.2%)	2	721 (0.82%)
			3	67 (0.077%)	3	35 (0.040%)
			4	2 (0.0023%)		
High	143349	2896647	0	92692 (65%)	0	100862 (71%)
			1	39450 (28%)	1	34519 (24%)
			2	8559 (6.0%)	2	5940 (4.2%)
			3	1208 (0.85%)	3	673 (0.47%)
			4	131 (0.092%)	4	58 (0.041%)
			5	20 (0.014%)	5	7 (0.0049%)
					6	1 (0.00070%)

Table 3.1: A summary of three mock catalogs. From left to right, it shows catalog type, observable BBH and BNS per year (note this is not detectable), and distributions of numbers of overlapping signals among BBH events. For example, in median merger rate catalog, there are 13125 detected BBH events (15% of all detected BBH events) coming with 1 detected overlapping GW signal, and 10461 detected BBHs coming with 1 undetected overlapping GW signal. The overlapping signal can be BBH or BNS, and two signals are defined as overlapped if their merger time difference $\Delta t < 4$s.

(NSBH) mergers and other possible types of sources; we use an analytical merger rate that peaks at $z \sim 2$, ignoring compact binaries from Pop III stars. Our catalogs aim to generate an appropriate merger rate for the study of systematic error accumulation, rather than accurately modeling the astrophysical population. To achieve this, we also adjust the merger rate to different levels, expecting that the real situation will lie somewhere between our lowest and highest estimates.

Signals are injected into the 3rd generation GW detector Einstein Telescope with ET-D PSD [166] located at the Cascina site of the current Virgo detector, and Cosmic Explorer located at the LIGO Hanford site with the sensitivity curve proposed by [174]. The frequency band used for the analysis is 5–2048 Hz.

3.2.2 Combining multiple events

There are several ways of combining results from multiple events [396, 397]. We employ two straightforward methods: multiplying likelihoods (equivalently, multiplying posteriors if priors are flat) and multiplying Bayes factors. The former assumes the modification parameter is the same for all events, while the latter allows the modification parameter to vary across events.

We assume a flat prior distribution, and that the posterior follows a multivariate Gaussian distribution with covariance matrix $\mathbf{\Gamma}^{-1}$ and mean $\boldsymbol{\mu}$ equal to injection values

$\theta_{\rm inj}$ plus systematic errors $\theta_{\rm sys}$. The statistical uncertainty of a parameter is $\sigma_i = \sqrt{(\mathbf{\Gamma}^{-1})_{ii}}$. The error ratio is then defined by Eq. 3.10.

In order to combine results from multiple events, one would multiply the posterior distributions of the testing parameter for each. Multiplication of Gaussian distributions results in another Gaussian distribution whose mean (systematic error) is a linear combination of the original means. From the first event in a catalog, we multiply the posterior of new events one by one and calculate the error ratio. Considering the arbitrary sequence of events, we permute the sequence 200 times and extract the ensemble average and 68% confidence interval.

Treating GR as a sub-model of the non-GR theory, Bayes factor can be calculated analytically with the Gaussian posterior [387]. Denote systematic error of $\delta\hat{\phi}_0$ as $\Delta\theta_{\rm sys}$, we have

$$L_{\rm GR}(\boldsymbol{\theta}_{\rm GR}) = L_{\rm nonGR}(\boldsymbol{\theta}_{\rm nonGR})\mid_{\delta\hat{\phi}_0=\Delta\theta_{\rm sys}}. \tag{3.12}$$

The Bayes factor is then calculated as

$$\begin{aligned}\mathcal{B}^{\rm nonGR}_{\rm GR} &\sim \frac{Z_{\rm nonGR}}{Z_{\rm GR}} = \frac{\int \mathbf{d}\boldsymbol{\theta}_{\rm nonGR} L_{\rm nonGR}}{\int \mathbf{d}\boldsymbol{\theta}_{\rm GR} L_{\rm GR}} \\ &= \sqrt{2\pi} e^{\frac{1}{2}(\mathbf{\Gamma}_{\delta\hat{\phi}_0\delta\hat{\phi}_0} - \mathbf{v}^{\rm T}(\mathbf{\Gamma}_{\rm GR})^{-1}\mathbf{v})\Delta\theta^2_{\rm sys}} \sqrt{\frac{\det \mathbf{\Gamma}_{\rm GR}}{\det \mathbf{\Gamma}_{\rm nonGR}}},\end{aligned} \tag{3.13}$$

where $\mathbf{\Gamma}_{\rm nonGR}$ is the Fisher matrix including the testing parameter, while $\mathbf{\Gamma}_{\rm GR}$ only includes GR parameters. $\mathbf{v}_i = (\partial h/\partial\boldsymbol{\theta}_i|\partial h/\partial\delta\hat{\phi}_0)$ represents the correlation between GR and non-GR parameters. $\mathbf{\Gamma}_{\delta\hat{\phi}_0\delta\hat{\phi}_0} = (\partial h/\partial\delta\hat{\phi}_0|\partial h/\partial\delta\hat{\phi}_0)$. The exponential term in the Bayes factor accounts for the deviation of GR, while the determinant ratio term usually favors GR since modified theories introduce extra parameters to explain the data. We also note that the correlation term $\mathbf{v}^{\rm T}(\mathbf{\Gamma}_{\rm GR})^{-1}\mathbf{v}$ mitigates the deviation of GR. Ignoring this term may overestimate the Bayes factor (e.g., [387]). When combining events, Bayes factors are numerically multiplied, with the same permutation used in multiplying likelihood. We consider a false deviation from GR to be achieved when $\ln \mathcal{B}^{\rm nonGR}_{\rm GR} > 8$. We reemphasize that Bayes factors are first computed for each event and then combined across the catalog, rather than calculated after different posteriors are multiplied. This analysis should be interpreted as not assuming that the testing parameter is the same for all events. In this sense it is less sensitive to violations of GR when there is a common underlying deviation parameter, so we would expect it to be less vulnerable to simulated false violations. While error ratio accumulation is decided by errors from each event, Bayes factor accumulation is more sensitive to the fraction of correct analyses in the catalog. The two methods of combining results are independent and do not necessarily lead to the same conclusion. More details are given in the next section.

3.2.3 Results

Overview of systematic errors in the catalog

We calculate the error ratio and Bayes factor for each BBH event in our mock catalog. In Fig. 3.3, we show the relation between SNR with the absolute error, error ratio and Bayes factor of the testing parameter $\delta\hat{\phi}_0$.

The results are consistent with the speculation that systematic error may dominate the high-SNR scenario. It is shown that the statistical error scales as 1/SNR, while the systematic errors roughly remain constant. The error ratio could exceed one for the "current" waveform, and this happens more often when SNR$>$ 30 despite the fact that high SNR events are rarer. Error ratios for the "future" waveform simulations are usually below one, but a certain amount of exceptions exist. For the Bayes factor analysis we find a similar situation, although there is a smaller fraction of more extreme values. There are roughly 0.8% and 18% events producing $\mathcal{R}(\delta\hat{\phi}_0) > 1$ for $\delta\hat{\beta}_2 = 5\times10^{-4}$ and 5×10^{-2} cases, respectively, while for $\ln\mathcal{B}^{\rm nonGR}_{\rm GR} > 8$ the fractions are 0.02% and 3%.

Error accumulation in a catalog

As pointed out by Moore et al. [387], false deviations could be achieved even though estimations for individual events are generally accurate. Here we combine all BBH events by multiplying likelihoods or Bayes factors and the results are shown Fig. 3.4. Let $N_{\rm event}$ be the number of events. When multiplying likelihoods, the statistical uncertainty shrinks as $1/\sqrt{N_{\rm event}}$. The absolute error of the testing parameter also decreases, but at a slower pace due to the perturbations from newly accumulating systematic errors. It also follows $1/\sqrt{N_{\rm event}}$ if there were no systematic errors - we observe that the test with the perfect waveform in a low merger rate catalog is approximately doing so. In most simulations, it is the waveform inaccuracy that keeps contributing to the systematic errors. The slower decay of systematic error results in a climbing error ratio as the number of events increases. At some point (typically $\sim 10^3$ events, considering error bars) it leads to a false deviation of GR for the "current" waveform. For the better waveform, the error ratio climbs as well, but it keeps below the statistical level until $10^5 - 10^6$ events.

Multiplying Bayes factors is a direct addition of $\ln\mathcal{B}^{\rm nonGR}_{\rm GR}$. "Correct analyses" can effectively decrease the combined Bayes factor so that a correct-analyses-dominated catalog leads to correct conclusions. Since there are only 3% of events with $\ln\mathcal{B}^{\rm nonGR}_{\rm GR} > 8$ (furthermore, only 7% of events with $\ln\mathcal{B}^{\rm nonGR}_{\rm GR} > 0$) for the current waveform, the sum of all Bayes factors is negative, thus false deviation is not achieved in this case. In contrast, multiplying likelihoods linearly adds systematic errors: for Gaussian distributions f and g, the mean of their product is $\mu_{fg} = \frac{\mu_f\sigma_g^2+\mu_g\sigma_f^2}{\sigma_f^2+\sigma_g^2}$. Correct analysis and different signs of errors could diminish systematic error a bit, but it is never guaranteed for the error to be

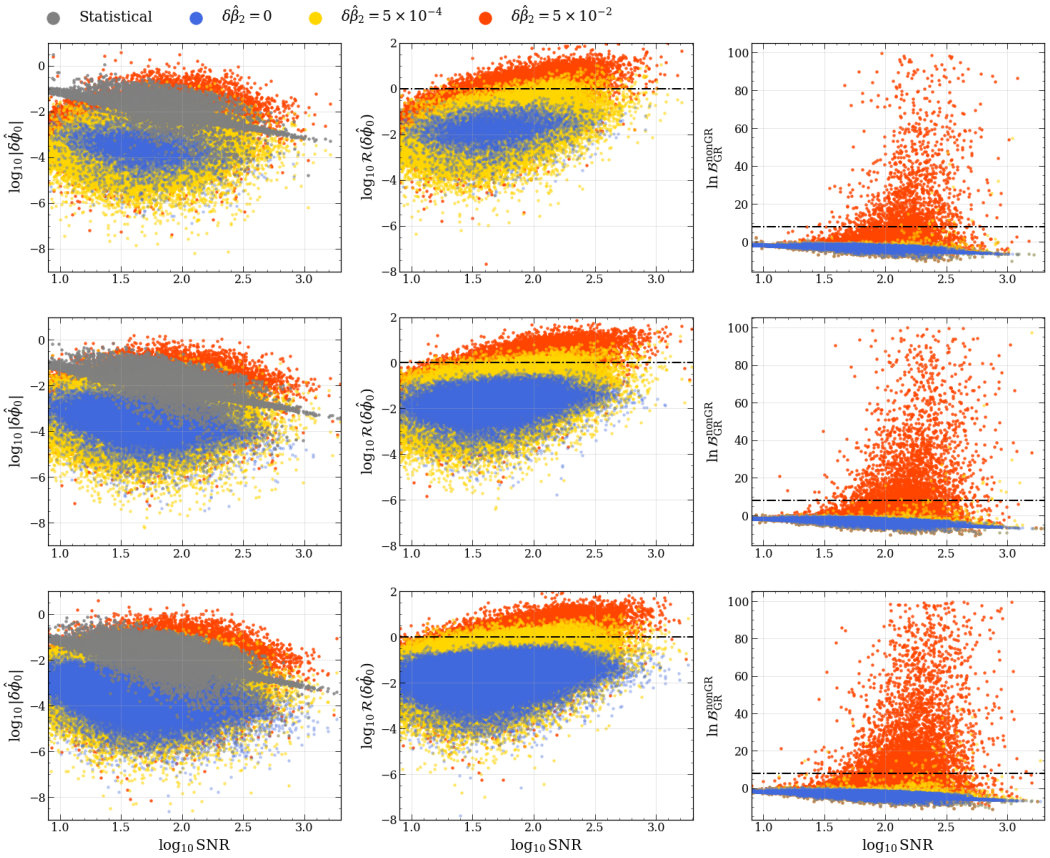


Figure 3.3: Relation between SNR and absolute error (first column), error ratio (second column) of $\delta\hat{\phi}_0$ and Bayes factor (third column) for low (uppermost row), median (median row), and high (bottom row) merger rate catalogs. Each point represents a BBH event. Blue points are for the "perfect waveform" case, where all systematic errors come from overlapping signals; red points stand for the "current waveform" case and yellow points for the "future waveform" case. Grey points in the first column are statistical errors. Dashed lines in the second and third columns are the threshold above which GR is mistakenly disfavoured. This plot shows $\mathcal{R}(\delta\hat{\phi}_0) > 1$ and $\ln \mathcal{B}^{\rm nonGR}_{\rm GR} > 8$ are mostly from "current waveform" and high SNR events.

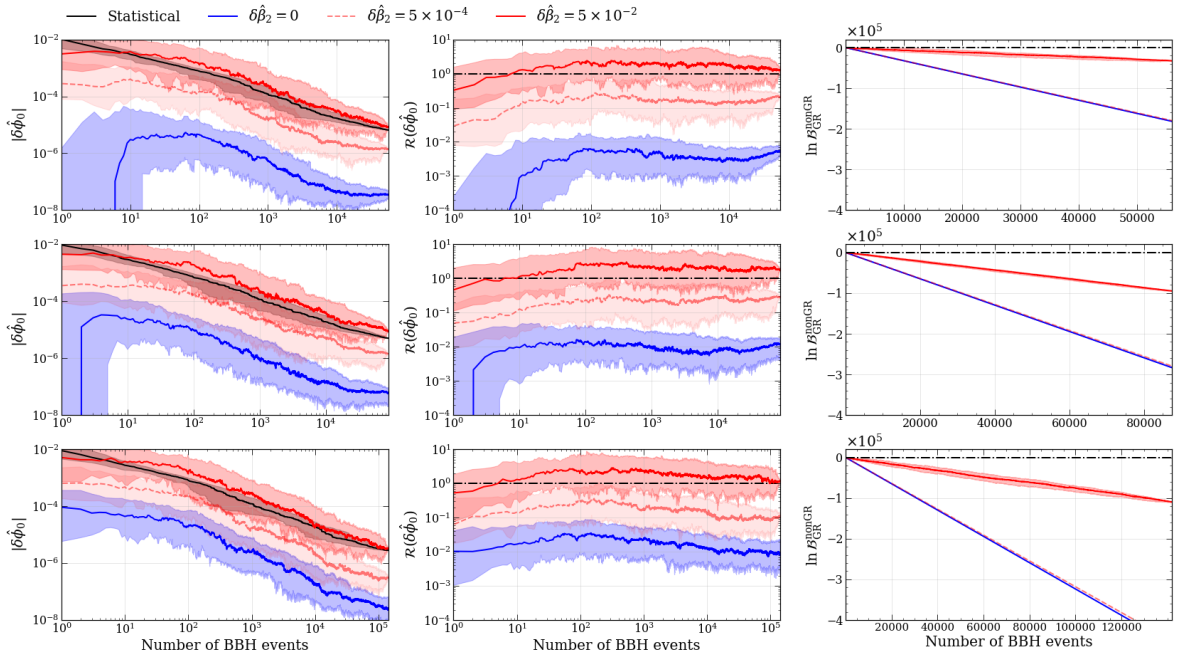


Figure 3.4: Systematic error accumulates with the increase of the number of events. The first column shows the absolute error of $\delta\hat{\phi}_0$ and the second column shows the error ratio. The third column is the Bayes factor. Solid red lines are the ensemble average for "current waveform", dashed red lines are for "future waveform", and blue lines stand for the perfect waveform. The shadow along lines is the 68% confidence interval. The first, second, and third rows are for low, median, and high catalogs, respectively. The black dotted-dashed line is the threshold above which a false deviation of GR is claimed. False deviations can be achieved with the increase in the number of events by multiplying posteriors, but multiplying Bayes factors does not give wrong conclusions in these correct-analyses-dominated catalogs.

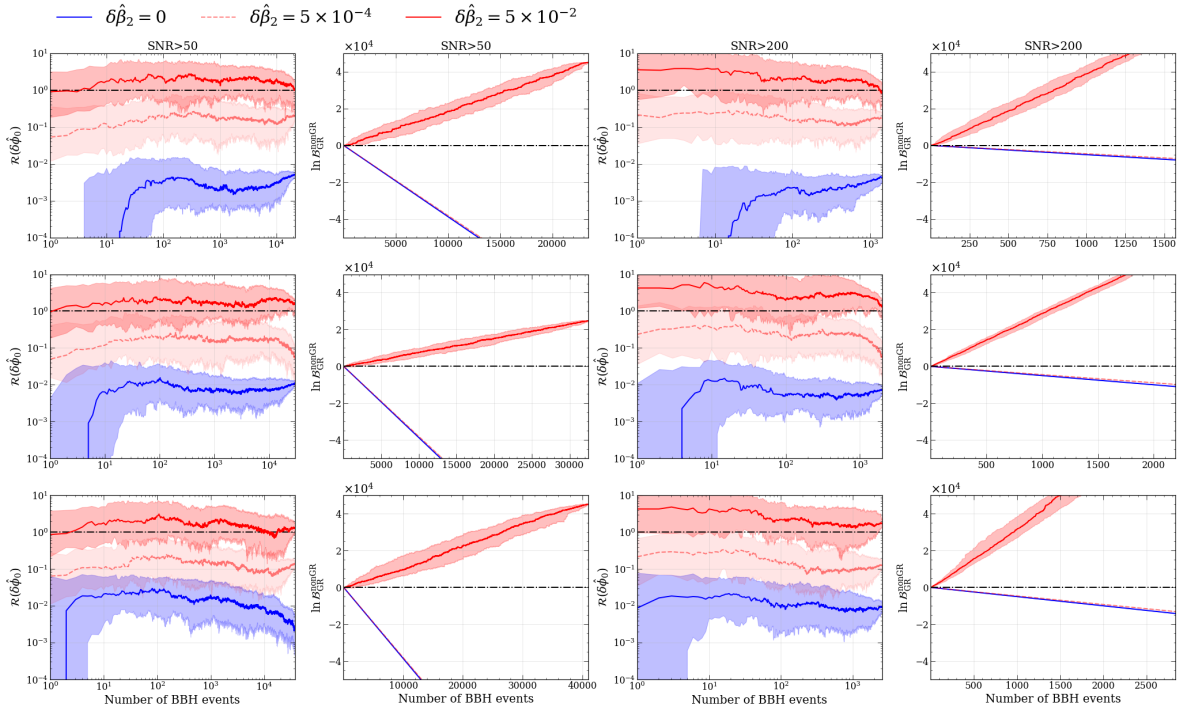


Figure 3.5: Similar to Fig. 3.4, the error ratio and Bayes factor accumulation. The left two columns show results from SNR>50 events, the right two columns are for SNR>200 events. Compared with Fig. 3.4, it shows that tests with high SNR events are more likely to make a false deviation from GR.

held around 0. Moreover, statistical uncertainty also shrinks during event stacking, so the error ratio shows a clear increase.

Golden events

We have combined all the detected BBH events in the above subsection. It is also interesting to test GR with only the "golden events", i.e., the GW events with high SNR and clean data that contribute to most of the information in the whole catalog test. This idea is widely used in many works, such as recent GWTC-3 tests of GR [25] and cosmology [152]. Since the noise is stationary and Gaussian in our simulation, we select the golden events with only two criteria: SNR above a chosen threshold (50 or 200) and there are no detected overlapping signals.

Results for the error ratio and Bayes factor are shown in Fig. 3.5: high SNR events are more vulnerable to systematic errors. Fewer events are needed to create a false deviation for the "current" waveform model, and the "future" waveform is closer to false deviation in all three catalogs. Moreover, the golden events catalog consists of more incorrect analyses ($\mathcal{R}(\delta\hat{\phi}_0) > 1$ or $\ln \mathcal{B}^{\rm nonGR}_{\rm GR} > 8$), and it causes the Bayes factor of current waveform to incorrectly favor the non-GR theory.

The statistical uncertainty decreases as 1/SNR while systematics do not as long as

the waveform is imperfect. The false deviation for golden events is not surprising from this angle, but it does need more attention and an appropriate solution for future data analysis.

Stacking residuals

Different types of systematic errors are correlated and could be a magnifying factor for each other, as expected from Eq. 3.3. With our mock catalog, we are able to investigate the effects of stacking residuals due to incorrect signal removals. In the high merger rate catalog, we calculate systematic errors for each BBH event, and show the distributions of systematic errors and Bayes factors caused by different numbers of overlapping signals in Fig. 3.6. It is shown that, with the increase in the number of overlapping signals, detected overlaps tend to produce larger errors, while errors from confusion background signals make smaller incremental changes. This verifies that overlapping signals could magnify the effects of waveform systematics, emphasizing the importance of waveform accuracy in the overlapping signal analysis.

3.3 Summary and discussions

We have investigated how systematic errors in testing GR accumulate under the influence of overlapping signals and inaccurate waveforms. We have considered different levels of waveform inaccuracies and event rates, and employed two approaches to combining the results.

We confirm that systematic errors could accumulate when combining multiple events, and could lead to incorrectly disfavoring GR in some cases. Since overlapping signals do not always occur, it is waveform inaccuracies that keep contributing to the systematic error in the catalog tests. An accurate waveform model is effective at preventing false deviations in most cases, while a worse one could lead to biased conclusions. We additionally find that overlapping signals can enlarge the effect of waveform systematics. By increasing the number of overlaps, we tend to achieve a greater systematic error and a Bayes factor that leans more toward the non-GR model. One can avoid this correlated error by selecting events with no detected overlapping signals, and, if one prefers, with high SNR as well. However, we have shown these events produce biases much faster because waveform systematics dominate in high SNR scenarios.

We should point out that GR is assumed to be the true theory to describe the data in this work, which is not necessarily correct. The inverse problem, namely, what happens to detection and PE when we use GR waveform for data analysis but GR is wrong (stealth bias), is investigated in previous works [369, 370, 398]. The core idea of our work and stealth bias is the same: using an incorrect model in data analysis can lead to biased

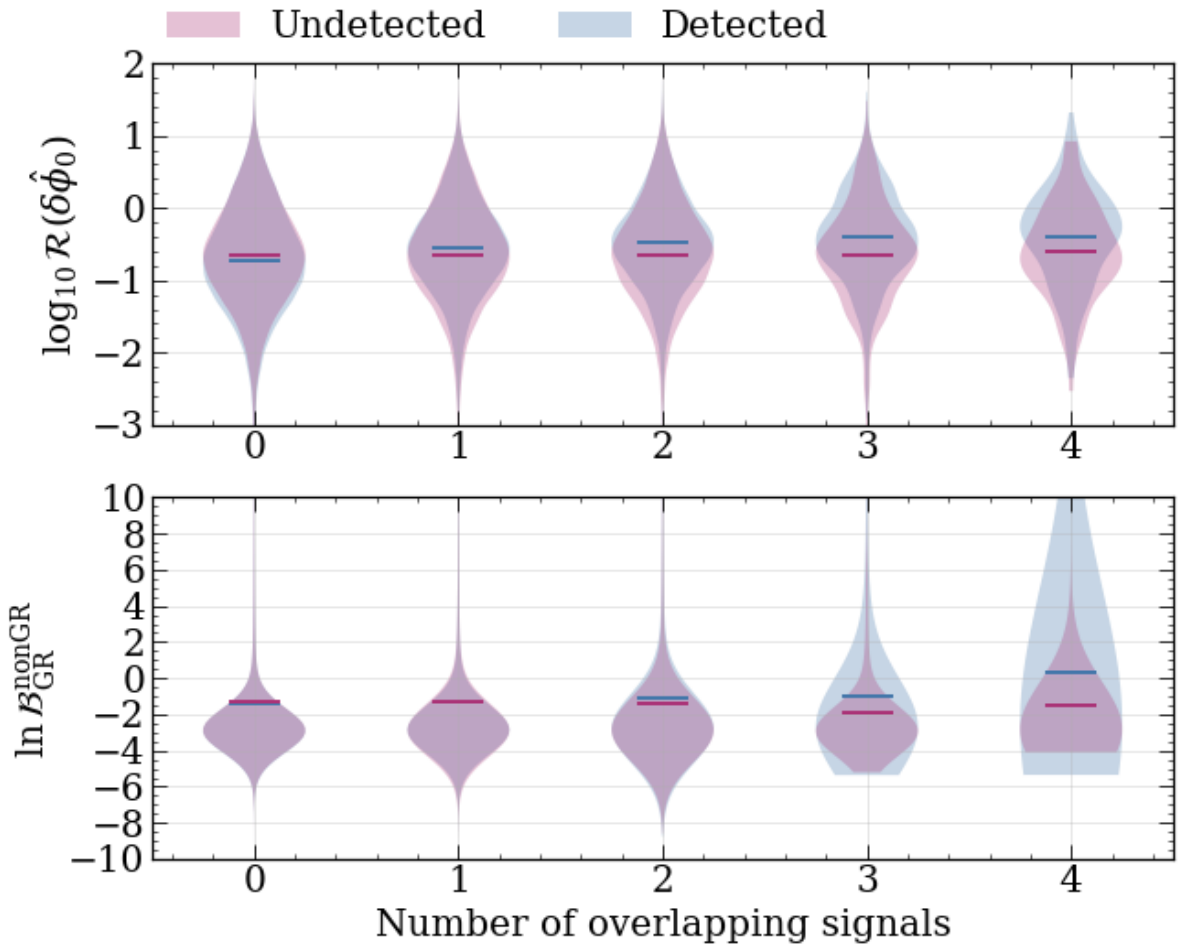


Figure 3.6: PPN test results for events with different overlapping signals. We show distributions of systematic errors and Bayes factors in the high merger rate catalog with $\delta\hat{\beta}_2 = 5 \times 10^{-2}$ waveform, classified by number and type of overlapping signals. Bars denote the mean value. The number of overlapping signals is cut at 4 because of the insufficient number of events coming with > 4 overlapping signals. The difference in the increase of mean values shows that the detected overlapping signals could magnify the effects of inaccurate waveform models.

results. Stealth bias emphasizes the importance of assuming the correct theory, while our work points out that even if the assumed fundamental theory is correct, waveform modeling and overlapping signals are still able to corrupt the results.

We re-emphasize that systematic errors can accumulate when combining multiple events and lead to incorrect scientific conclusions. This problem is universal: in addition to tests of GR, any analysis based on a GW catalog is faced with this issue, such as constraints on cosmological models, neutron star models [386], and astrophysical population inference. Furthermore, there are more sources of systematic errors than those investigated in this work: instrumental calibration [399, 400], glitches [401, 402], missing physical effects [278, 279] and so forth. A recent review [403] summarized the potential causes of false deviations from GR. A full analysis of these contributions, and their relative importance, will be essential in designing analysis strategies for 3G detectors. An obvious solution to these issues is continuing improvements to waveform model accuracy and instrumental stability, but we believe more efforts are needed from the angle of data analysis. A proper estimate of confusion background may be necessary [385], and new techniques might be needed, such as accounting for waveform systematic errors during PE [266], performing specific analysis of residual strain [404], and so forth.

We focus on the catalog level in this chapter, combining multiple events in which some show deviations from GR while some do not. However, we will pay attention to every single event that shows deviations when analyzing real data, which requires a careful assessment of systematic errors in each event. As waveform inaccuracies contribute most to systematic errors, in the next chapter, I will introduce a method of assessing waveform accuracy for parameter estimation and its potential applications.

Chapter 4

Assessing the accuracy of waveform models

As we have seen in Chapter 3, with the improvement in the sensitivity of GW detectors and the increasing diversity of GW sources, there is a strong need for accurate GW waveform models for data analysis. While the current model accuracy assessments require waveforms generated by Numerical Relativity (NR) simulations as the "true waveforms", in this chapter we propose an assessment approach that does not require NR simulations, which enables us to probe into model accuracy everywhere in the parameter space. By measuring the difference between two waveform models, we derive a necessary condition for a pair of waveform models to both be accurate, for a particular set of parameters. We then apply this method to the parameter estimation samples of the GWTC-2.1 and GWTC-3, and find that the waveform accuracy for high signal-to-noise ratio events in some cases fails our assessment criterion. Based on the analysis of real events' posterior samples, we discuss the correlation between our quantified accuracy assessments and systematic errors in parameter estimation. We find waveform models that perform worse in our assessment are more likely to give inconsistent estimations. We also investigate waveform accuracy in different parameter regions, and find the accuracy degrades as the spin effects go up, the mass ratio deviates from one, or the orbital plane is near-aligned to the line of sight. Furthermore, we make predictions of waveform accuracy requirements for future detectors and find the accuracy of current waveform models should be improved by at least 3 orders of magnitude, which is consistent with previous works.

This chapter is organized as follows. We introduce the methods of assessing waveform accuracy in Sec. 4.1 and apply them on BBH waveforms in Sec. 4.2 and BNS and NSBH waveforms in Sec. 4.3. A summary and discussions about future applications of this method will be given in Sec. 4.4.

4.1 Methods of assessing waveform accuracy

4.1.1 Assessing waveform accuracy with NR simulations

We have shown that an accurate waveform is crucial to correct data analysis, especially in high-SNR scenarios. Ideally, the waveform model accuracy should be estimated when it is used to analyze a GW event. Waveform models are usually calibrated and validated against accurate NR simulations, and the mismatch (Eq. 1.151) is used to quantify the accuracy. However, the mismatch is not relevant to the SNR because it is normalized by the *norms* of waveforms. Here the norm of a waveform is defined as

$$\|h\| = \sqrt{(h|h)}, \tag{4.1}$$

and can be understood as the "length" of strains in the noise-weighted inner product space. We want to quantify whether a waveform model is accurate enough for the analysis, therefore the SNR needs to be accounted for.

A model waveform can be thought as "accurate enough" when the detector can not distinguish it from the real one. Ref. [405] constructs a waveform family H to quantify the detector's ability to measure the difference between the model and the real waveform: Let h_0 be the true waveform, h_1 be the waveform given model, and $\delta h_1 = h_1 - h_0$ represents their difference. We construct the following waveform family of the model

$$H_1(\lambda) = (1-\lambda)h_0 + \lambda h_1 = h_0 + \lambda \delta h_1, \quad 0 < \lambda < 1, \tag{4.2}$$

where λ is a parameter which interpolates between the two models. If the measurement error on λ is greater than the length of its domain of definition (i.e. the parametric distance between real and model waveforms), we can claim the detector is not able to distinguish the waveforms, thus the model is accurate enough. The error σ_λ is given by the Fisher matrix formalism

$$\sigma_\lambda^{-2} = \left(\frac{\partial H_1}{\partial \lambda} \middle| \frac{\partial H_1}{\partial \lambda}\right) = (\delta h_1 \mid \delta h_1). \tag{4.3}$$

Therefore, the accuracy standard for a waveform model is

$$\|\delta h_1\|^2 = (\delta h_1 \mid \delta h_1) < 1. \tag{4.4}$$

Eq. 4.4 implies the waveform difference should lie within a unit sphere in the inner product space, any violation of which means the model is not accurate enough. Since $\langle n|n\rangle = 1$, another way to understand Eq. 4.4 is that if the distance to the real waveform is longer than the length of detector noise, the detector will be able to tell the error of the model. From this angle, the waveform we are considering here should be the detector response,

i.e.,

$$\begin{aligned} h_0 &= F_+ h_0^+ + F_\times h_0^\times \\ h_1 &= F_+ h_1^+ + F_\times h_1^\times, \end{aligned} \tag{4.5}$$

where $F_+, F_\times$ are antenna response functions. Let $\rho = \|h_0\| \approx \|h_1\|$ be the SNR of the signal, and MM be the mismatch between h_0 and h_1 defined by Eq. 1.151, Eq. 4.4 can be written as

$$\rho < \sqrt{\frac{1}{2\mathrm{MM}}}, \tag{4.6}$$

which gives the SNR threshold for assessing waveform accuracy. Given a waveform model and its mismatch with the true waveform, there could be potential systematic errors if it is used to analyze signals that are above the SNR threshold. We note that some works [405, 406, 407] propose a less stringent criterion than Eq. 4.4 by changing the upper limit to $2\epsilon\rho^2$, where ρ is SNR and ϵ is the maximum tolerated fractional loss in SNR which needs to be appropriately chosen for detection. In this chapter, we focus more on the waveform systematics in measurement rather than in detection, so we keep using Eq. 4.4, i.e., the strict distinguishability criterion.

However, to compute $\|\delta h_1\|$, the true waveform h_0 is needed, which is usually replaced by the computationally expensive NR simulations that can not span all over the parameter space. It is therefore impractical to apply Eq. 4.4 directly in data analysis, and the waveform uncertainties are unknown if the source parameters are out of the model's calibration range. This may cause some unknown systematic errors in data analysis. An example is GW191219_163120 [128], of which mass ratio is estimated to be out of the waveform calibration region (≤ 0.041) so that there are potential uncertainties in its p_{astro}. We need a NR-free waveform accuracy criteria that can be performed in the entire parameter space.

4.1.2 NR-free waveform accuracy criteria

Assessment of the detector response

It is easy to verify that the Cauchy-Schwarz inequality and the triangle inequality hold for the noise-weighted inner product space:

$$\|a\|^2\|b\|^2 \geq |(a|b)|^2 \tag{4.7}$$

$$\|a\| + \|b\| \geq \|a \pm b\| \geq |\ \|a\| - \|b\|\ |\ . \tag{4.8}$$

Therefore, instead of using one waveform model and the true waveform, we can introduce another waveform model h_2 to be paired with h_1. Although δh_1 and δh_2 are unknown,

their difference $\delta h_1 - \delta h_2$ can be easily calculated:

$$\begin{aligned}\Delta &= \delta h_1 - \delta h_2 \\ &= (h_1 - h_0) - (h_2 - h_0) \\ &= h_1 - h_2.\end{aligned} \tag{4.9}$$

Assuming both of two waveforms are accurate, i.e., they both satisfy Eq. 4.4, we can obtain an upper limit of $\|\Delta\|$ using the triangle inequality:

$$\|\Delta\| \leq \|\delta h_1\| + \|\delta h_2\| < 2. \tag{4.10}$$

Eq. 4.10 is a necessary condition if h_1 and h_2 are both accurate. Note that the threshold of 2 is built on the threshold of 1 for each waveform model, corresponding to the 1-sigma error. However, due to the magnification of the triangle inequality, the threshold of 2 should have a confidence level greater than 1-sigma. That is to say, if Eq. 4.10 is violated, at least one of the waveform models does not satisfy Eq. 4.4 with confidence level greater than 1-sigma.

We illustrate possible cases for the $\|\Delta\|$ in Eq. 4.10 in the vector plots Fig. 4.1, in which waveforms are treated as vectors in the noise-weighted inner product space. The black circle denotes the sphere of radius 2. Vectors δh_1 and δh_2 denote the difference between the real waveform h_0 and the models h_1, h_2, respectively, and different line styles denote different possibilities. δh_i lies in the circle means the ith model is accurate and satisfies Eq. 4.4. If the length of Δ is greater than the upper limit 2 (the diameter of the black circle), as shown in case I, *at least* one of the waveform model errors can not be put inside the circle, i.e., it does not meet the accuracy standard. However, $\|\Delta\| < 2$ does not mean both of the waveforms are accurate, as shown in Case II. Small $\|\Delta\|$ only implies the two models give similar predictions of the waveform but can not guarantee their accuracy. The key idea of this method is: if two waveforms have significant differences, they can not both be correct.

Eq. 4.10 is not a strong criterion; it can not tell which waveform causes the violation (case I), and may miss some waveform errors (case II). Despite this, we suppose it still gives certain information about the correctness of waveform modeling. If $\|\Delta\| > 2$, the waveform pair should become less reliable; if $\|\Delta\| \gg 2$, the systematic errors in waveform models should not be neglected as it is highly possible that either of the waveforms is accurate, or one of them has seriously deviated. If $\|\Delta\| < 2$, no evidence of waveform inaccuracy is found by this approach, although we could not exclude the possibility that two waveforms have large but similar errors. The advantage of this method is that it can be performed everywhere in the parameter space, as long as waveform models work in that region.

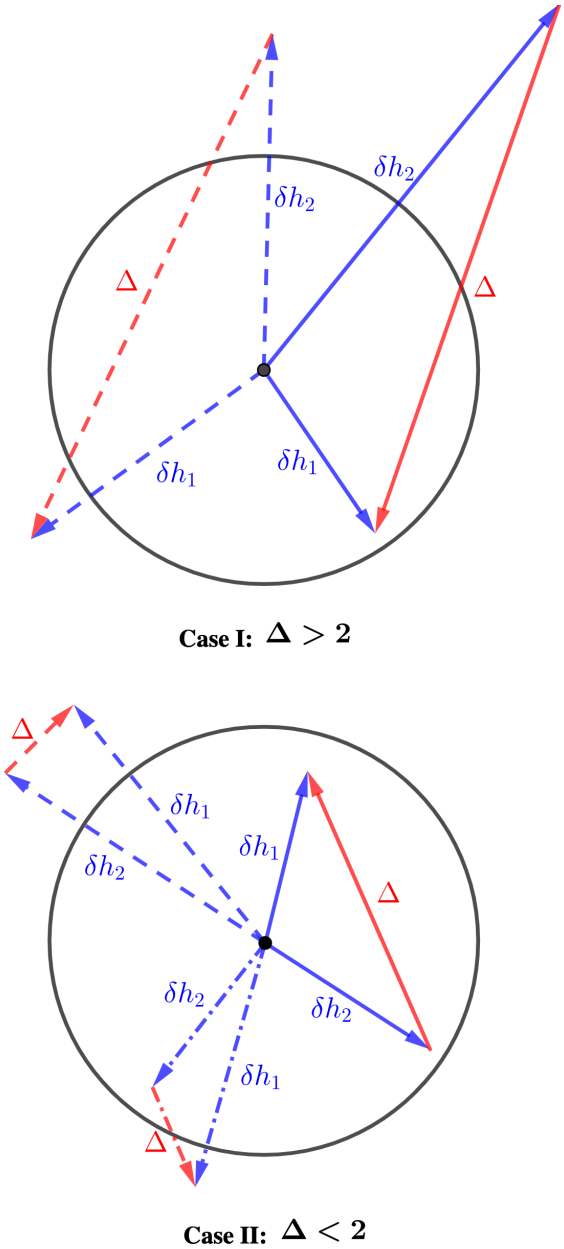


Figure 4.1: Vector plots to illustrate all cases of Δ. Blue vectors are the difference between waveform models and the real waveform, and black circles represent the sphere of radius 2, the upper limit of length of δh_i if h_i is accurate ($i = 1, 2$). Red vectors are Δ, the difference between two waveform models (defined in Eq. 4.9). Different line styles denote different possibilities. In Case I, $\|\Delta\|$ exceeds the upper limit given by Eq. 4.10, so at least one in h_1, h_2 is not accurate enough. In Case II, $\|\Delta\|$ satisfies Eq. 4.10, there may be 0, 1, 2 inaccurate waveforms, corresponding to solid line, dotted-dashed line and dashed line, respectively. We can not determine the accuracy of a waveform pair in Case II.

Eq. 4.10 can be extended to a detector network by defining the inner product between matrices by Eq. 2.25. The signal of the network can be denoted as a column vector $\mathbf{h} = (h^{(1)}, h^{(2)}, \ldots, h^{(\mathrm{N_d})})^{\mathrm{T}}$, where superscript (k) denotes the k-th detector and $\mathrm{N_d}$ is the number of detectors in the network. We can also subtract two waveform models, and define $\mathbf{h_1} - \mathbf{h_2} = \Delta_{\mathrm{net}}$. The norm of Δ_{net} can be calculated

$$\begin{aligned} \|\Delta_{\mathrm{net}}\|^2 &= (\delta\mathbf{h}^{\mathbf{T}}|\delta\mathbf{h}) = \sum_k (\delta h^{(k)}|\delta h^{(k)}) \\ &= \sum_k \left(\Delta^{(k)}\right)^2 < 4\mathrm{N_d}, \end{aligned} \tag{4.11}$$

where $F_+^{(k)}, F_\times^{(k)}$ are the antenna response functions of the k-th detector. In practice, we can weight the Δ by the number of detectors:

$$\Delta'_{\mathrm{net}} = \frac{\Delta_{\mathrm{net}}}{\sqrt{\mathrm{N_d}}}, \tag{4.12}$$

so that the Δ'_{net} will have an upper limit of 2 if the waveforms are both accurate enough.

At last, we should note that the inner product in the calculation of waveform difference Δ should be minimized over an arbitrary phase ϕ_0 and time shift t_0 as described in Sec. 1.4.3 in order to eliminate the kinematical difference between models.

Assessment of the detector response

The two accuracy standards we proposed, Eq. 4.10 and Eq. 4.12, are related to the SNR, as the length of Δ is proportional to the amplitude of GWs. It is reasonable that the higher the SNR is, the easier it is for detectors to distinguish different waveforms, and the more important systematic errors will be in data analysis. However, SNR depends on not only intrinsic parameters, but also extrinsic parameters that trivially modulate the amplitude. It is the phase evolution that is critical to reveal the physical properties of the source, and is the intrinsic characteristic of a GW waveform [408]. We therefore normalize the Δ with SNR to eliminate the impacts from amplitudes. The optimal SNR is defined as $\rho = \sqrt{(h|h)}$, which is also proportional to the amplitude of GWs like Δ. Thus we have $\Delta \propto \rho$. In fact, we have two waveforms to calculate Δ. The normalization factor is chosen as the geometric mean of SNRs from two waveforms, i.e., $\rho_0 = \sqrt{\rho_1\rho_2}$. Take Eq. 4.10 as an example, the normalized Δ is

$$\begin{aligned} \|\Delta_{\mathrm{SNR}=1}\|^2 &= \frac{(\delta h_1 - \delta h_2|\delta h_1 - \delta h_2)}{\sqrt{(h_1|h_1)(h_2|h_2)}} \\ &= \frac{(h_1 - h_2|h_1 - h_2)}{\sqrt{(h_1|h_1)(h_2|h_2)}}, \end{aligned} \tag{4.13}$$

and we simply have

$$\|\Delta_{\mathrm{SNR}=\rho_0}\| = \rho_0 \|\Delta_{\mathrm{SNR}=1}\|. \tag{4.14}$$

Eq. 4.14 can be used to evaluate waveform accuracy at a fixed SNR. Note the threshold of $\|\Delta_{\mathrm{SNR}=\rho_0}\|$ is always 2. The normalized $\|\Delta\|$ can be related to the mismatch by

$$\|\Delta_{\mathrm{SNR}=1}\|^2 \approx 2\mathrm{MM}, \tag{4.15}$$

if we assume $\rho_1 \approx \rho_2$. However, we found ρ_1 and ρ_2 could have considerable differences in some parameter regions when comparing `IMRPhenomXPHM` and `SEOBNRv4PHM`, and this will be discussed in Sec. 4.2.

In the next two sections, we will apply the accuracy standard Eq. 4.12 and Eq. 4.14 to GW waveforms from 3 types of compact binary coalescence: BBH, NSBH, and BNS. We employ the assessment on the GWTC parameter estimation samples and parameter grids we generate; the former aims to investigate whether faulty waveforms were used in GW data analysis and possible systematic error caused by waveform errors, while the latter explores waveforms' performances in different regions of the parameter space. Throughout this chapter, we ignore the calibration error, which can cause our waveforms to be slightly different from those used in GWTC-3 and GWTC-2.1 parameter estimation. The calibration error is typically < 4 degrees in phase and $< 7\%$ in amplitude [399], and it acts on both waveform models, so ignoring it will not have large impacts on our results.

4.2 Applications on BBH waveforms

BBH mergers are the most frequent GW events at this stage: Among all 91 GW candidates (36 in GWTC-3 [128], 44 in GWTC-2.1 [127] and 11 in GWTC-1 [125]), over 80 of them are confirmed to be BBH events. In the latest data release from LVK collaboration, waveform models `IMRPhenomXPHM` and `SEOBNRv4PHM` are used for analysis of all the BBH events, including re-analysis of GWTC-1 events published in GWTC-2.1. Due to the low SNR of current NSBH events, the resolution of tidal deformability is poor and no strong sign of matter effects is revealed in data analysis. Besides, higher modes and spin precession effects are more important than matter effects for waveform modeling of NSBHs [409], so `IMRPhenomXPHM` and `SEOBNRv4PHM` are also employed on NSBH events to extract physical information.

For all the 89 BBH and NSBH events, we use the cosmologically reweighted parameter estimation posterior samples from GWTC-3 and GWTC-2.1 data release and calculate $\|\Delta'_{\mathrm{net}}\|$ (Eq. 4.12) of the waveform models mentioned above. We use the mixture of `IMRPhenomXPHM` and `SEOBNRv4PHM` samples in most events, but in some events `SEOBNRv4PHM` samples are not provided [127], so we use `IMRPhenomXPHM` samples to calcu-

late $\|\Delta'_{\rm net}\|$ between `IMRPhenomXPHM` and `SEOBNRv4PHM`. Samples we use are the same as GWTC-3 [128] and GWTC-2.1 [127]. For each sample, we generate the waveform (including the detector response) for both models, then apply a time and phase shift on one of them to minimize Eq. 4.12. The minimized $\|\Delta'_{\rm net}\|$ is the waveform difference we refer to in the following discussion.

When $\|\Delta'_{\rm net}\|$ is greater than 2 at a sampling point, it implies the difference between `IMRPhenomXPHM` and `SEOBNRv4PHM` is so large at this point that they could not both be accurate enough. Furthermore, the difference in waveform will induce a difference in likelihood, and therefore has the potential to affect the results of a parameter estimation algorithm. This yields a systematic difference in parameter estimates, and so the results from different waveform models may not coincide. Therefore, in addition to $\|\Delta'_{\rm net}\|$, we also calculate Jensen-Shannon (J-S) divergence between `IMRPhenomXPHM` samples and `SEOBNRv4PHM` samples (if available). The J-S divergence is a measurement of the similarity between two probability distributions and is used in GWTC-2 [126]. The greater it is, the greater the difference between the two distributions and there may be potential systematic errors in the data analysis. Since the J-S divergence for samples from a distribution is easiest to evaluate in one dimension, we choose the greatest J-S divergence among samples for the following parameters: mass ratio q, chirp mass $\mathcal{M}$, effective spin $\chi_{\rm eff}$ and effective precession spin $\chi_{\rm p}$ as a measurement of inconsistency of posterior samples, for they are the major physical parameters to be studied. We use `gaussian_kde` in `SciPy` to estimate probability density functions. The base of J-S divergence is chosen to be 2, so that the divergence ranges between 0 and 1.

The full results of the 89 BBH and NSBH events are shown in Tab. 4.1 and Tab. 4.2 for reference. We list the basic information of each event, including some source parameters and network SNR, and statistics we construct, including the mean value of $\|\Delta'_{\rm net}\|$, normalized $\|\Delta'_{\rm net}\|$ (which equals to $\|\Delta'_{\rm net}\|/\mathrm{SNR}$), the fraction of $\|\Delta'_{\rm net}\| < 2$ samples, and the J-S divergence. We highlight some points in the rest of this subsection.

4.2.1 Overall accuracy

We show the relations between the waveform difference $\|\Delta'_{\rm net}\|$ of different events and SNR in Fig. 4.2, in which each point represents a GW event. We find every event has samples that can not meet the $\|\Delta'_{\rm net}\| < 2$ requirement (left panel), but most events have mean $\|\Delta'_{\rm net}\|$ around 2 (right panel). This means some waveform pairs used in data analysis can pass (and are near the edge of) our accuracy standard, but violations exist. We could not identify whether one or both waveform models is inaccurate. Later in Sec. 4.2.3 we will show it is the samples with large spin or small mass ratio or edge-on inclination that contribute to $\|\Delta'_{\rm net}\| < 2$ fraction. Overall, considering that the violations are generally not strong, we conclude that the current waveform accuracy is around the edge of our

Event name	m_1	m_2	$\chi_{\rm eff}$	$\chi_{\rm p}$	$\theta_{\rm JN}$	SNR	Mean	$\rm Mean_{Norm}$	Fraction	Max J-S Div.
GW150914_095045	$37.99^{+5.11}_{-2.89}$	$32.96^{+3.18}_{-5.03}$	$-0.04^{+0.12}_{-0.14}$	$0.51^{+0.35}_{-0.38}$	$2.70^{+0.32}_{-0.71}$	24.4	2.0	0.08	0.59	0.069
GW151012_095443	$29.62^{+17.08}_{-6.87}$	$16.30^{+5.73}_{-6.02}$	$0.12^{+0.28}_{-0.21}$	$0.36^{+0.43}_{-0.27}$	$1.70^{+1.13}_{-1.40}$	10	1.33	0.13	0.81	0.023
GW151226_033853	$15.50^{+12.19}_{-4.00}$	$8.18^{+2.62}_{-3.02}$	$0.20^{+0.23}_{-0.08}$	$0.52^{+0.36}_{-0.35}$	$0.88^{+2.00}_{-0.66}$	13.1	2.39	0.18	0.54	N/A
GW170104_101158	$34.78^{+7.73}_{-4.85}$	$25.25^{+4.43}_{-5.60}$	$-0.04^{+0.15}_{-0.19}$	$0.40^{+0.40}_{-0.31}$	$1.10^{+1.79}_{-0.86}$	13	1.28	0.1	0.88	0.01
GW170608_020116	$11.41^{+4.36}_{-1.49}$	$8.38^{+1.21}_{-2.06}$	$0.05^{+0.13}_{-0.05}$	$0.32^{+0.41}_{-0.24}$	$2.37^{+0.58}_{-2.06}$	14.9	1.43	0.1	0.86	0.051
GW170729_185629	$77.88^{+17.89}_{-14.48}$	$44.13^{+18.11}_{-17.46}$	$0.29^{+0.25}_{-0.33}$	$0.39^{+0.40}_{-0.29}$	$1.35^{+1.44}_{-1.03}$	10.8	2.08	0.19	0.54	0.037
GW170809_082821	$41.19^{+9.63}_{-6.27}$	$29.25^{+5.63}_{-6.46}$	$0.07^{+0.17}_{-0.17}$	$0.39^{+0.44}_{-0.30}$	$2.61^{+0.39}_{-0.59}$	12.4	1.14	0.09	0.91	0.037
GW170814_103043	$34.70^{+6.01}_{-3.52}$	$28.03^{+3.22}_{-4.59}$	$0.08^{+0.13}_{-0.12}$	$0.48^{+0.38}_{-0.37}$	$0.69^{+1.92}_{-0.48}$	15.9	1.31	0.08	0.88	0.062
GW170818_022509	$42.14^{+7.99}_{-5.04}$	$33.42^{+5.07}_{-6.14}$	$-0.06^{+0.19}_{-0.22}$	$0.56^{+0.34}_{-0.41}$	$2.46^{+0.47}_{-0.50}$	11.3	1.06	0.09	0.96	0.015
GW170823_131358	$51.67^{+11.92}_{-7.85}$	$39.42^{+8.11}_{-11.03}$	$0.05^{+0.21}_{-0.22}$	$0.47^{+0.41}_{-0.35}$	$1.73^{+1.16}_{-1.48}$	11.5	1.83	0.16	0.64	0.019
GW190408_181802	$31.70^{+6.96}_{-3.88}$	$23.83^{+3.56}_{-5.01}$	$-0.03^{+0.13}_{-0.17}$	$0.37^{+0.41}_{-0.29}$	$1.01^{+1.85}_{-0.79}$	14.4	1.05	0.07	0.94	0.006
GW190412_053044	$31.76^{+6.81}_{-6.60}$	$10.34^{+2.19}_{-1.55}$	$0.21^{+0.12}_{-0.13}$	$0.19^{+0.22}_{-0.12}$	$0.92^{+1.69}_{-0.40}$	18.2	2.86	0.16	0.25	0.45
GW190413_052954	$52.79^{+15.00}_{-10.02}$	$37.90^{+9.95}_{-11.57}$	$-0.04^{+0.27}_{-0.32}$	$0.44^{+0.42}_{-0.33}$	$0.79^{+2.01}_{-0.58}$	8.5	1.13	0.13	0.88	N/A
GW190413_134308	$83.12^{+19.30}_{-15.64}$	$50.15^{+19.47}_{-23.80}$	$-0.01^{+0.28}_{-0.38}$	$0.55^{+0.36}_{-0.41}$	$1.85^{+1.01}_{-1.52}$	10.3	1.57	0.15	0.71	N/A
GW190421_213856	$60.79^{+13.06}_{-8.71}$	$46.94^{+9.04}_{-14.55}$	$-0.10^{+0.21}_{-0.27}$	$0.45^{+0.41}_{-0.34}$	$2.03^{+0.85}_{-1.70}$	9.7	1.59	0.16	0.73	N/A
GW190503_185404	$53.32^{+12.17}_{-10.28}$	$36.45^{+10.17}_{-12.62}$	$-0.05^{+0.23}_{-0.30}$	$0.43^{+0.40}_{-0.33}$	$2.50^{+0.46}_{-0.62}$	12.2	1.55	0.13	0.74	0.013
GW190512_180714	$29.34^{+6.90}_{-6.72}$	$15.83^{+4.64}_{-2.97}$	$0.02^{+0.13}_{-0.14}$	$0.26^{+0.41}_{-0.20}$	$1.79^{+1.07}_{-1.50}$	12.2	1.3	0.11	0.85	0.036
GW190513_205428	$49.99^{+14.75}_{-12.78}$	$25.60^{+10.88}_{-6.98}$	$0.16^{+0.29}_{-0.22}$	$0.35^{+0.43}_{-0.26}$	$0.79^{+2.01}_{-0.58}$	12.3	1.64	0.13	0.7	0.048
GW190514_065416	$66.96^{+21.39}_{-12.57}$	$47.33^{+12.11}_{-18.23}$	$-0.08^{+0.29}_{-0.35}$	$0.45^{+0.42}_{-0.33}$	$1.47^{+1.38}_{-1.20}$	8.3	1.44	0.17	0.78	0.013
GW190517_055101	$52.92^{+14.76}_{-9.99}$	$32.75^{+9.55}_{-12.25}$	$0.49^{+0.21}_{-0.28}$	$0.55^{+0.31}_{-0.32}$	$2.12^{+0.70}_{-1.18}$	10.8	2.04	0.19	0.61	0.141
GW190519_153544	$94.80^{+15.84}_{-12.40}$	$59.88^{+16.88}_{-18.85}$	$0.33^{+0.20}_{-0.24}$	$0.45^{+0.36}_{-0.28}$	$1.61^{+0.95}_{-1.02}$	12.4	2.81	0.23	0.24	0.066
GW190521_030229	$152.36^{+31.65}_{-17.62}$	$89.65^{+48.96}_{-52.17}$	$-0.14^{+0.50}_{-0.45}$	$0.49^{+0.33}_{-0.35}$	$1.38^{+1.40}_{-1.07}$	13.3	3.24	0.24	0.35	N/A
GW190521_074359	$52.19^{+7.65}_{-5.39}$	$40.36^{+5.87}_{-7.16}$	$0.10^{+0.13}_{-0.13}$	$0.39^{+0.37}_{-0.30}$	$1.48^{+1.27}_{-1.10}$	24.4	3.6	0.15	0.13	0.153
GW190527_092055	$51.05^{+29.88}_{-9.27}$	$32.36^{+11.95}_{-13.36}$	$0.10^{+0.22}_{-0.22}$	$0.36^{+0.47}_{-0.28}$	$1.15^{+1.65}_{-0.87}$	8.7	1.29	0.15	0.85	0.161
GW190602_175927	$106.81^{+24.90}_{-17.99}$	$67.63^{+23.03}_{-31.56}$	$0.12^{+0.25}_{-0.28}$	$0.45^{+0.43}_{-0.34}$	$2.13^{+0.79}_{-1.87}$	12.3	1.95	0.16	0.6	N/A
GW190620_030421	$87.13^{+23.93}_{-17.28}$	$53.35^{+17.64}_{-24.30}$	$0.34^{+0.22}_{-0.29}$	$0.48^{+0.38}_{-0.32}$	$2.07^{+0.82}_{-1.74}$	10.9	2.1	0.19	0.52	0.049
GW190630_185205	$41.40^{+8.12}_{-6.50}$	$28.24^{+5.79}_{-5.63}$	$0.10^{+0.14}_{-0.13}$	$0.33^{+0.36}_{-0.24}$	$1.41^{+1.43}_{-1.16}$	15.2	1.55	0.1	0.76	0.032
GW190701_203306	$74.56^{+15.98}_{-10.85}$	$56.08^{+12.08}_{-17.93}$	$-0.08^{+0.23}_{-0.31}$	$0.44^{+0.41}_{-0.33}$	$0.58^{+0.55}_{-0.42}$	11.7	0.95	0.08	0.95	0.007
GW190706_222641	$117.53^{+22.57}_{-18.55}$	$63.74^{+27.20}_{-27.42}$	$0.28^{+0.25}_{-0.31}$	$0.47^{+0.38}_{-0.33}$	$1.86^{+0.96}_{-1.50}$	12.5	3.06	0.24	0.26	0.093
GW190707_093326	$14.07^{+3.01}_{-2.33}$	$9.28^{+1.69}_{-1.48}$	$-0.04^{+0.10}_{-0.09}$	$0.28^{+0.39}_{-0.22}$	$2.12^{+0.81}_{-1.89}$	13.2	1.23	0.09	0.93	0.219
GW190708_232457	$23.41^{+4.97}_{-5.07}$	$13.70^{+3.57}_{-2.17}$	$0.05^{+0.10}_{-0.10}$	$0.26^{+0.44}_{-0.20}$	$1.39^{+1.54}_{-1.18}$	13.1	1.14	0.09	0.94	0.424
GW190719_215514	$58.83^{+75.12}_{-16.34}$	$32.78^{+14.38}_{-16.38}$	$0.25^{+0.33}_{-0.32}$	$0.45^{+0.39}_{-0.32}$	$1.61^{+1.26}_{-1.33}$	8	1.41	0.18	0.79	N/A
GW190720_000836	$16.57^{+7.36}_{-3.94}$	$8.76^{+2.44}_{-2.10}$	$0.19^{+0.14}_{-0.11}$	$0.29^{+0.39}_{-0.20}$	$2.59^{+0.41}_{-1.99}$	11.5	0.86	0.08	0.97	0.14
GW190727_060333	$58.87^{+13.07}_{-8.01}$	$46.18^{+8.22}_{-13.24}$	$0.09^{+0.25}_{-0.27}$	$0.50^{+0.38}_{-0.37}$	$1.54^{+1.35}_{-1.27}$	12.1	1.37	0.11	0.82	0.017
GW190728_064510	$14.56^{+8.19}_{-2.60}$	$9.39^{+1.92}_{-2.91}$	$0.13^{+0.19}_{-0.07}$	$0.29^{+0.39}_{-0.20}$	$1.11^{+1.77}_{-0.90}$	13.4	1.16	0.09	0.89	0.048
GW190731_140936	$64.91^{+15.63}_{-10.80}$	$46.27^{+12.67}_{-17.54}$	$0.07^{+0.28}_{-0.25}$	$0.41^{+0.43}_{-0.32}$	$1.23^{+1.63}_{-0.98}$	8.5	1.13	0.13	0.89	0.031
GW190803_022701	$57.71^{+13.23}_{-8.92}$	$42.96^{+9.20}_{-13.81}$	$-0.01^{+0.23}_{-0.28}$	$0.44^{+0.42}_{-0.34}$	$0.91^{+1.93}_{-0.70}$	9.1	0.87	0.1	0.96	N/A
GW190814_211039	$24.48^{+1.55}_{-1.43}$	$2.72^{+0.11}_{-0.11}$	$0.00^{+0.07}_{-0.07}$	$0.04^{+0.04}_{-0.03}$	$0.90^{+1.40}_{-0.24}$	22.2	2.54	0.11	0.49	N/A
GW190828_063405	$43.27^{+7.36}_{-4.65}$	$35.40^{+4.69}_{-6.98}$	$0.15^{+0.15}_{-0.16}$	$0.43^{+0.41}_{-0.32}$	$2.38^{+0.57}_{-2.02}$	16.3	1.31	0.08	0.86	N/A
GW190828_065509	$30.42^{+8.31}_{-8.29}$	$13.44^{+4.82}_{-2.71}$	$0.05^{+0.16}_{-0.17}$	$0.26^{+0.42}_{-0.20}$	$1.86^{+0.96}_{-1.52}$	11.1	1.69	0.15	0.68	N/A
GW190910_112807	$56.54^{+8.69}_{-6.47}$	$44.50^{+7.47}_{-9.24}$	$-0.00^{+0.17}_{-0.20}$	$0.38^{+0.43}_{-0.30}$	$1.62^{+1.16}_{-1.25}$	13.4	2.25	0.17	0.47	0.011
GW190915_235702	$43.05^{+11.11}_{-6.00}$	$32.50^{+5.62}_{-7.95}$	$-0.03^{+0.19}_{-0.24}$	$0.56^{+0.34}_{-0.40}$	$1.84^{+0.99}_{-1.48}$	13	1.63	0.13	0.69	0.01
GW190924_021846	$9.78^{+4.83}_{-2.00}$	$5.67^{+1.35}_{-1.61}$	$0.03^{+0.20}_{-0.08}$	$0.25^{+0.41}_{-0.19}$	$0.84^{+1.95}_{-0.64}$	13	1.13	0.09	0.92	0.152
GW190929_012149	$101.70^{+25.99}_{-18.32}$	$41.64^{+25.15}_{-18.10}$	$-0.03^{+0.23}_{-0.28}$	$0.31^{+0.51}_{-0.25}$	$1.45^{+1.18}_{-0.97}$	10.1	2.22	0.22	0.47	N/A
GW190930_133541	$16.36^{+9.27}_{-4.57}$	$7.98^{+2.78}_{-2.34}$	$0.19^{+0.22}_{-0.16}$	$0.30^{+0.42}_{-0.21}$	$0.72^{+2.06}_{-0.55}$	10.1	1.2	0.12	0.87	0.377

Table 4.1: The first half of the BBH analysis results, including 10 GWTC-1 events and 35 GWTC-2.1 events that are included in GWTC-2. First seven columns are basic information of the events: event names in YYMMDD_HHMMSS form, component masses $m_{1,2}$ in detector frame (which have a difference of factor $1+z$ from Ref. [127, 128], z is the cosmological redshift), effective spin $\chi_{\rm eff}$, effective precession spin $\chi_{\rm p}$, inclination angle $\theta_{\rm JN}$ and network SNR. The parameters are shown by 50% percentile and 90% confidence error bar, but note $\theta_{\rm JN}$ usually has a bimodal distribution, the 50% percentile might be misleading. **Last four columns are statistics we construct**: mean value and the normalized mean value of $\|\Delta'_{\rm net}\|$ among all the samples (the latter one is simply the mean value divided by network SNR), fraction of $\|\Delta'_{\rm net}\| < 2$ samples, and the maximum J-S divergence between samples of q, $\mathcal{M}$, $\chi_{\rm eff}$ and $\chi_{\rm p}$. It is labeled as "N/A" if the parameter estimation result from one of the waveforms is not included in GWTC-2.1 data release.

Event name	m_1	m_2	$\chi_{\rm eff}$	$\chi_{\rm p}$	$\theta_{\rm JN}$	SNR	Mean	$\text{Mean}_{\text{Norm}}$	Fraction	Max J-S Div.
GW190403_051519	$185.98^{+37.48}_{-58.13}$	$44.64^{+61.34}_{-24.08}$	$0.68^{+0.16}_{-0.43}$	$0.32^{+0.38}_{-0.22}$	$1.84^{+1.12}_{-1.66}$	8	1.62	0.2	0.72	0.052
GW190426_190642	$178.48^{+87.83}_{-31.54}$	$132.65^{+32.32}_{-63.94}$	$0.23^{+0.42}_{-0.41}$	$0.51^{+0.37}_{-0.36}$	$2.08^{+0.80}_{-1.72}$	9.6	1.55	0.16	0.74	N/A
GW190725_174728	$14.23^{+12.21}_{-3.62}$	$7.59^{+2.40}_{-3.00}$	$-0.04^{+0.36}_{-0.16}$	$0.37^{+0.46}_{-0.28}$	$1.00^{+1.79}_{-0.74}$	9.1	1.04	0.11	0.92	N/A
GW190805_211137	$87.19^{+24.88}_{-15.28}$	$59.60^{+16.40}_{-23.31}$	$0.37^{+0.29}_{-0.39}$	$0.50^{+0.34}_{-0.32}$	$1.00^{+1.75}_{-0.74}$	8.3	1.3	0.16	0.84	0.014
GW190916_200658	$78.32^{+32.30}_{-21.54}$	$42.37^{+24.07}_{-21.18}$	$0.20^{+0.33}_{-0.31}$	$0.37^{+0.43}_{-0.28}$	$1.61^{+1.27}_{-1.35}$	8.2	1.14	0.14	0.89	0.004
GW190917_114630	$11.15^{+3.72}_{-4.48}$	$2.35^{+1.21}_{-0.48}$	$-0.08^{+0.21}_{-0.43}$	$0.17^{+0.42}_{-0.13}$	$1.35^{+1.60}_{-1.16}$	9.5	1.06	0.11	0.93	N/A
GW190925_232845	$24.69^{+7.70}_{-3.17}$	$18.46^{+2.68}_{-4.16}$	$0.09^{+0.16}_{-0.15}$	$0.39^{+0.43}_{-0.30}$	$0.77^{+2.07}_{-0.58}$	9.9	1.09	0.11	0.91	0.008
GW190926_050336	$63.55^{+31.78}_{-14.11}$	$31.96^{+21.57}_{-14.17}$	$-0.02^{+0.25}_{-0.33}$	$0.37^{+0.48}_{-0.29}$	$1.67^{+1.03}_{-1.19}$	9	1.49	0.17	0.74	N/A
GW191103_012549	$14.03^{+7.42}_{-2.34}$	$9.42^{+1.79}_{-2.85}$	$0.21^{+0.16}_{-0.10}$	$0.40^{+0.41}_{-0.26}$	$1.38^{+1.52}_{-1.14}$	8.9	1.01	0.11	0.92	0.012
GW191105_143521	$13.00^{+4.54}_{-1.78}$	$9.36^{+1.45}_{-2.19}$	$-0.02^{+0.13}_{-0.09}$	$0.30^{+0.45}_{-0.24}$	$1.07^{+1.82}_{-0.85}$	9.7	0.77	0.08	0.99	0.02
GW191109_010717	$81.16^{+12.89}_{-8.89}$	$59.72^{+15.58}_{-17.43}$	$-0.29^{+0.42}_{-0.31}$	$0.63^{+0.29}_{-0.37}$	$1.91^{+0.87}_{-1.18}$	17.3	5.8	0.34	0.1	0.086
GW191113_071753	$36.10^{+14.71}_{-16.18}$	$7.31^{+6.49}_{-1.57}$	$0.00^{+0.37}_{-0.29}$	$0.20^{+0.54}_{-0.16}$	$1.70^{+1.08}_{-1.32}$	7.9	1.68	0.21	0.67	0.048
GW191126_115259	$15.71^{+7.24}_{-2.51}$	$10.75^{+1.94}_{-2.98}$	$0.21^{+0.15}_{-0.11}$	$0.39^{+0.40}_{-0.26}$	$1.71^{+1.20}_{-1.48}$	8.3	1.14	0.14	0.89	0.01
GW191127_050227	$86.41^{+60.12}_{-37.31}$	$38.45^{+31.09}_{-25.26}$	$0.18^{+0.34}_{-0.36}$	$0.52^{+0.41}_{-0.41}$	$1.46^{+1.40}_{-1.16}$	9.2	1.9	0.21	0.6	0.089
GW191129_134029	$12.29^{+4.87}_{-2.26}$	$7.80^{+1.67}_{-1.94}$	$0.06^{+0.16}_{-0.08}$	$0.26^{+0.36}_{-0.19}$	$1.73^{+1.16}_{-1.46}$	13.1	1.34	0.1	0.87	0.033
GW191204_110529	$36.20^{+15.49}_{-5.66}$	$26.21^{+5.17}_{-7.54}$	$0.05^{+0.26}_{-0.27}$	$0.52^{+0.38}_{-0.39}$	$1.57^{+1.24}_{-1.24}$	8.8	1.59	0.18	0.72	0.027
GW191204_171526	$13.44^{+3.77}_{-1.98}$	$9.29^{+1.54}_{-1.84}$	$0.16^{+0.08}_{-0.05}$	$0.39^{+0.35}_{-0.26}$	$2.26^{+0.66}_{-2.00}$	17.5	1.67	0.1	0.73	0.045
GW191215_223052	$33.48^{+9.29}_{-4.68}$	$24.46^{+4.07}_{-5.17}$	$-0.04^{+0.17}_{-0.21}$	$0.50^{+0.37}_{-0.38}$	$1.20^{+1.50}_{-0.85}$	11.2	1.06	0.09	0.95	0.01
GW191216_213338	$12.95^{+4.92}_{-2.38}$	$8.23^{+1.73}_{-1.99}$	$0.11^{+0.13}_{-0.06}$	$0.23^{+0.35}_{-0.16}$	$2.50^{+0.44}_{-0.81}$	18.6	1.93	0.1	0.62	0.06
GW191219_163120	$34.73^{+2.27}_{-2.68}$	$1.30^{+0.08}_{-0.05}$	$-0.00^{+0.07}_{-0.09}$	$0.09^{+0.07}_{-0.07}$	$1.76^{+1.13}_{-1.49}$	9.1	2.51	0.28	0.34	0.13
GW191222_033537	$67.15^{+14.73}_{-9.54}$	$52.41^{+10.67}_{-15.40}$	$-0.04^{+0.20}_{-0.25}$	$0.41^{+0.41}_{-0.32}$	$1.62^{+1.24}_{-1.33}$	12.5	2.02	0.16	0.57	0.017
GW191230_180458	$82.70^{+19.46}_{-13.10}$	$62.85^{+13.88}_{-21.42}$	$-0.05^{+0.26}_{-0.31}$	$0.52^{+0.38}_{-0.39}$	$2.03^{+0.85}_{-1.67}$	10.4	1.17	0.11	0.88	0.011
GW200105_162426	$9.57^{+1.85}_{-1.82}$	$2.02^{+0.35}_{-0.25}$	$0.00^{+0.13}_{-0.18}$	$0.09^{+0.17}_{-0.07}$	$1.54^{+1.28}_{-1.22}$	13.7	1.47	0.11	0.79	0.162
GW200112_155838	$44.01^{+8.22}_{-5.16}$	$35.18^{+5.11}_{-7.42}$	$0.06^{+0.15}_{-0.15}$	$0.39^{+0.39}_{-0.30}$	$0.88^{+2.04}_{-0.68}$	19.8	1.63	0.08	0.73	0.061
GW200115_042309	$6.30^{+2.15}_{-2.69}$	$1.53^{+0.91}_{-0.30}$	$-0.15^{+0.24}_{-0.42}$	$0.20^{+0.34}_{-0.16}$	$0.62^{+1.94}_{-0.43}$	11.3	1.35	0.12	0.81	0.096
GW200128_022011	$65.05^{+15.99}_{-9.41}$	$51.16^{+10.32}_{-13.35}$	$0.12^{+0.24}_{-0.25}$	$0.57^{+0.34}_{-0.40}$	$1.38^{+1.46}_{-1.06}$	10.6	2.12	0.2	0.55	0.09
GW200129_065458	$40.25^{+12.23}_{-3.33}$	$34.06^{+3.32}_{-10.82}$	$0.11^{+0.11}_{-0.16}$	$0.52^{+0.42}_{-0.37}$	$0.66^{+0.59}_{-0.41}$	26.8	2.33	0.09	0.47	0.425
GW200202_154313	$11.02^{+3.83}_{-1.51}$	$7.99^{+1.21}_{-1.85}$	$0.04^{+0.13}_{-0.06}$	$0.28^{+0.40}_{-0.22}$	$2.57^{+0.42}_{-0.59}$	10.8	0.74	0.07	0.99	0.025
GW200208_130117	$52.95^{+12.23}_{-8.40}$	$38.51^{+8.64}_{-11.18}$	$-0.07^{+0.22}_{-0.27}$	$0.38^{+0.41}_{-0.29}$	$2.53^{+0.44}_{-0.58}$	10.8	1.19	0.11	0.87	0.017
GW200208_222617	$83.38^{+171.77}_{-48.68}$	$21.91^{+13.04}_{-11.81}$	$0.45^{+0.43}_{-0.44}$	$0.41^{+0.37}_{-0.30}$	$1.54^{+1.30}_{-1.22}$	7.4	1.91	0.26	0.6	0.171
GW200209_085452	$55.70^{+14.96}_{-9.88}$	$42.95^{+11.00}_{-13.60}$	$-0.12^{+0.24}_{-0.30}$	$0.51^{+0.39}_{-0.37}$	$1.74^{+1.12}_{-1.44}$	9.6	1.29	0.13	0.83	0.014
GW200210_092254	$28.68^{+8.45}_{-5.24}$	$3.38^{+0.52}_{-0.52}$	$0.02^{+0.22}_{-0.21}$	$0.15^{+0.22}_{-0.12}$	$2.31^{+0.60}_{-1.97}$	8.4	1.2	0.14	0.88	0.06
GW200216_220804	$84.36^{+28.39}_{-21.07}$	$50.74^{+22.50}_{-30.85}$	$0.10^{+0.34}_{-0.36}$	$0.45^{+0.42}_{-0.35}$	$0.89^{+1.87}_{-0.69}$	8.1	1.32	0.16	0.82	0.006
GW200219_094415	$58.57^{+13.46}_{-8.95}$	$44.41^{+9.26}_{-14.14}$	$-0.08^{+0.23}_{-0.29}$	$0.48^{+0.40}_{-0.35}$	$1.19^{+1.59}_{-0.92}$	10.7	1.61	0.15	0.71	0.017
GW200220_061928	$165.56^{+62.24}_{-27.93}$	$120.51^{+29.47}_{-55.27}$	$0.06^{+0.40}_{-0.38}$	$0.50^{+0.37}_{-0.37}$	$1.24^{+1.55}_{-0.96}$	7.2	1.2	0.17	0.87	0.022
GW200220_124850	$64.13^{+16.51}_{-10.34}$	$46.89^{+11.54}_{-16.61}$	$-0.07^{+0.27}_{-0.33}$	$0.49^{+0.39}_{-0.37}$	$1.66^{+1.17}_{-1.34}$	8.5	1.72	0.2	0.68	0.007
GW200224_222234	$52.30^{+9.08}_{-5.37}$	$42.82^{+5.83}_{-9.89}$	$0.10^{+0.15}_{-0.15}$	$0.49^{+0.37}_{-0.36}$	$0.62^{+0.55}_{-0.45}$	20	2.26	0.11	0.52	0.022
GW200225_060421	$23.58^{+5.65}_{-3.25}$	$17.20^{+3.05}_{-4.55}$	$-0.12^{+0.17}_{-0.28}$	$0.53^{+0.34}_{-0.38}$	$1.31^{+1.47}_{-1.00}$	12.5	1.43	0.11	0.8	0.015
GW200302_015811	$48.45^{+10.33}_{-9.58}$	$25.50^{+12.02}_{-7.63}$	$0.01^{+0.25}_{-0.26}$	$0.37^{+0.45}_{-0.28}$	$1.34^{+1.43}_{-1.01}$	10.8	2.13	0.2	0.51	0.054
GW200306_093714	$39.52^{+21.47}_{-10.80}$	$20.98^{+8.86}_{-9.91}$	$0.32^{+0.28}_{-0.46}$	$0.43^{+0.39}_{-0.31}$	$1.12^{+1.74}_{-0.87}$	7.8	1.33	0.17	0.82	0.013
GW200308_173609	$143.98^{+299.28}_{-85.60}$	$52.82^{+83.32}_{-33.18}$	$0.16^{+0.58}_{-0.49}$	$0.41^{+0.42}_{-0.30}$	$1.55^{+1.19}_{-1.27}$	7.1	1.1	0.15	0.89	0.054
GW200311_115853	$41.83^{+8.21}_{-4.59}$	$34.00^{+4.68}_{-7.26}$	$-0.02^{+0.16}_{-0.20}$	$0.45^{+0.40}_{-0.35}$	$0.55^{+0.52}_{-0.40}$	17.8	1.27	0.07	0.87	0.026
GW200316_215756	$15.98^{+12.23}_{-3.33}$	$9.52^{+2.32}_{-3.49}$	$0.13^{+0.27}_{-0.10}$	$0.29^{+0.38}_{-0.20}$	$2.32^{+0.58}_{-1.85}$	10.3	1.08	0.1	0.9	0.082
GW200322_091133	$105.87^{+392.13}_{-84.00}$	$26.23^{+58.87}_{-18.59}$	$0.08^{+0.51}_{-0.47}$	$0.50^{+0.36}_{-0.41}$	$1.66^{+1.04}_{-1.12}$	6	0.87	0.14	0.94	0.059

Table 4.2: The second half of the BBH analysis results, including 8 new events in GWTC-2.1 (compared to GWTC-2) and 36 events in GWTC-3. Columns have the same meaning as Tab. 4.1.

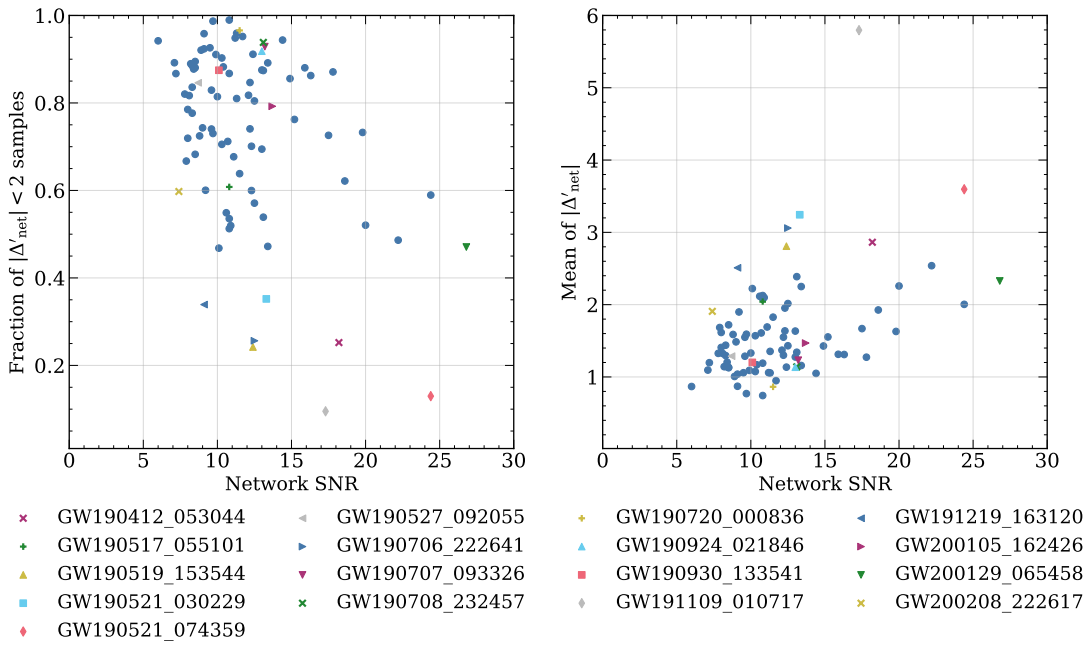


Figure 4.2: The left panel shows the relation between fraction of samples that meet our accuracy standard ($\|\Delta'_{\rm net}\| < 2$) and network SNR, and the right panel shows mean value of all samples' $\|\Delta'_{\rm net}\|$ and network SNR. We highlight the events whose $\|\Delta'_{\rm net}\| < 2$ sample fraction is less than 0.4 and whose maximum J-S divergence is greater than 0.1. These two plots show waveforms of higher SNR events are more likely to violate our waveform accuracy standard, and given the current detector sensitivity, we are already observing some events that violate our assessment criterion. Note the normalized $\|\Delta'_{\rm net}\|$ can also be read out from the right panel: it is the slope of the line that connects the origin and each point. We can compare the waveform difference of these events when they have the same SNR by comparing the slope. The numerical values are given in the 7th-10th columns of Tab. 4.1 and 4.2.

assessment standard for the current detector sensitivity which makes detections of SNRs ranging from 8 to ~ 30.

Although the properties of GW sources differ, there is a tendency that large SNR events are more likely to have greater waveform difference (as expected by Eq. 4.14), and have fewer samples that meet the $\|\Delta'_{\rm net}\| < 2$ requirement. This emphasizes the importance of waveform modeling for future GW detections, in which the SNR can reach hundreds to thousands. We can also make a rough estimation of waveform accuracy requirements for future detectors. The mismatch $\mathcal{M}$ with the "true" waveform is widely-used to assess the waveform accuracy, and the relation between $\|\Delta\|$ and $\mathcal{M}$ can be derived with Eqs. 4.14 and 4.15:

$$\mathrm{MM}(h_1, h_2) \approx \frac{1}{2\rho_0^2}\|\Delta_{\mathrm{SNR}=\rho_0}(h_1, h_2)\|^2 \tag{4.16}$$

Eq. 4.16 gives the mismatch between two waveform models, but limited by the triangle inequality, the mismatch between models $\mathcal{M}(h_1, h_2)$ should be at the same order of

magnitude as the mismatch between a model and the real waveform $\mathcal{M}(h_1, h_0)$, under the assumption that both models are well-calibrated by high precision waveforms like NR simulation. From our previous discussion, we know the $\|\Delta\|$ is around the edge of its upper limit under the current detector sensitivity. If we assume $\|\Delta\|$ is of the same range for future detectors, and SNR is roughly 30-100 times higher, we can determine that the mismatch should decrease 3-4 orders of magnitude. This is consistent with the results reported in Ref. [382].

4.2.2 Impacts on parameter estimation

From previous discussions, the waveforms generated by posterior samples of GWTC-3 and GTWC-2.1 are mostly within the waveform difference bound, yet there are some exceptions. Seven GW events have more than 60% posterior samples violating the standard, which means the difference of two waveform models might be too large to ensure their accuracies. The difference of waveforms may result in difference in parameter estimation, indicating systematic errors [368, 382, 410].

We show the relation between waveform difference and posterior sample consistency (maximum J-S divergence) in Fig. 4.3, where we can see a weak tendency that events with large waveform difference are more likely to have large J-S divergence, i.e., the difference in waveform models may lead to inconsistency in parameter estimation. Particularly, when the fraction of $\|\Delta'_{\rm net}\| < 2$ samples is below 40%, the maximum J-S divergence would be greater than the majority of the GW events. This coincides with our expectations.

However, the inverse statement is not necessarily true. When most posterior samples meet our accuracy standard, it is also possible that two waveform models give inconsistent results. In fact, the waveform error is not the only factor that causes two sets of posterior samples to differ. The behavior of samplers or packages (`bilby` [109] vs `RIFT` [411]) and the prior choice (such as high-spin and low-spin prior for neutron stars [136]) can both influence the consistency between the two posterior samples, although the latter one is not involved in our analysis. Even if we exclude these factors in a full Bayesian analysis, theoretically, it is the combination of waveform gradients, covariance matrices and waveform difference that contributes to systematic errors in parameter estimation [368], not just waveform difference. Besides, we use the maximum J-S divergence as the measurement of posterior difference, which might be influenced when the parameter estimation does not work efficiently on some specific parameters. The last row in Fig. 4.3 shows such cases. This makes the correlation between posterior sample consistency and waveform difference more statistically dispersed.

In the last three rows of Fig. 4.3, we give some examples of inconsistent posterior samples. GW200129_065458 has the largest J-S divergence among GWTC-3 events, and GW190412_053044 has the largest J-S divergence among GWTC-2.1 events. The pos-

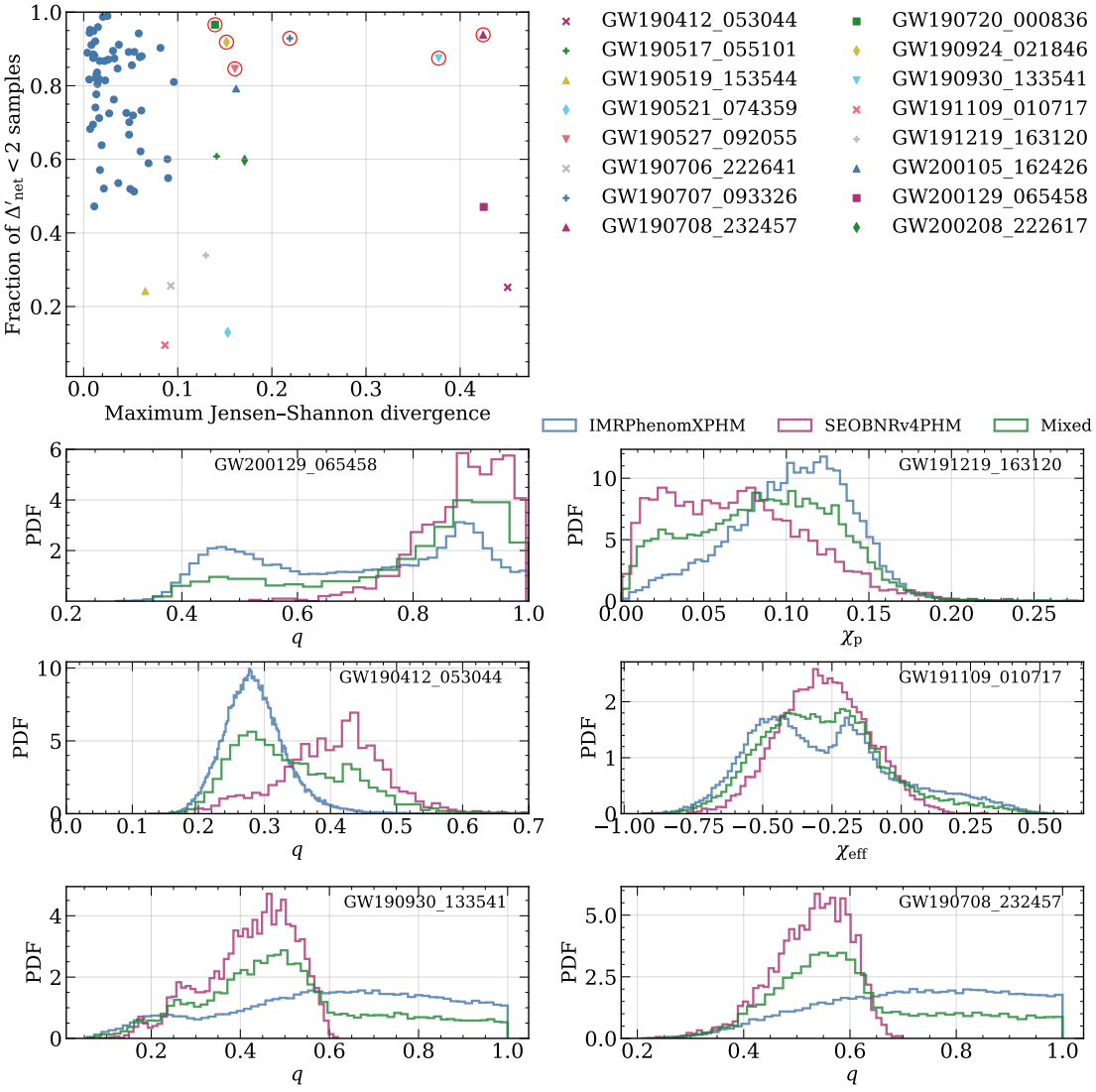


Figure 4.3: First row: We visualize the fraction of samples that meet our accuracy standard and maximum J-S divergence in $\{q, \mathcal{M}, \chi_{\rm eff}, \chi_{\rm p}\}$ of the two samples (10th and 11th columns of Tab. 4.1 and 4.2). We highlight the events whose fraction of $\|\Delta'_{\rm net}\| < 2$ samples is less than 40%, and the events whose maximum J-S divergence is greater than 0.1. Some GWTC-2.1 events have nearly flat `IMRPhenomXPHM` posteriors for mass ratio (as showed in the undermost row), which caused large J-S divergence despite the small waveform difference. We use red circles to label these events.
Bottom three rows: We show some examples of inconsistent posterior samples; the parameter name and event name are shown in the plots.

terior sample inconsistencies of the two events are also reported in GWTC-3 [128] and GWTC-2.1 [127]. In both events, the result with `IMRPhenomXPHM` suggests the possibility of a low mass ratio binary, while that with `SEOBNRv4PHM` does not. GW191219_163120 is the lowest mass ratio detection to date. Its estimated mass ratio is out of the calibration range of waveform models, so potential systematic error may lie in its data analysis [128]. In our analysis, GW191219_163120 does have fewer posterior samples that pass our assessment than most other events, but it is not the worst one. Besides, its high SNR (26.8) also contributes to waveform difference: its waveform difference becomes small after normalization. This might be caused by the small spins indicated by parameter estimation. Therefore, we suppose the waveform modeling is not that problematic in the low mass ratio and small spin region, but its high SNR reduces model waveform accuracy. We show its estimation of effective precession spin in Fig. 4.3, in which we see the result of `IMRPhenomXPHM` supports high precession effects in this binary system, while the result of `SEOBNRv4PHM` prefers lower precession effects. GW191109_010717 produces the largest waveform difference in our analysis. In a later section 4.2.3 we will illustrate it might be caused by its special spin effects and higher modes. We show its estimation of effective spin in Fig. 4.3: results from two waveform models show different multimodality. We also give examples that do not significantly violate our accuracy standard but have inconsistent posterior samples: GW190930_133541 and GW190708_232457. Their results from `IMRPhenomXPHM` seem unable to find the most probable mass ratio. There are six events having this behaviour in GWTC-2.1, as labeled by red circles in Fig. 4.3. Further investigation is needed, but this is beyond the scope of this work.

Since most posterior samples in this analysis satisfy or just slightly violate our accuracy standard, and samples from two waveform models, generated by different samplers and packages, are mixed as the final results to counterbalance systematic errors, we suppose the waveform modeling error will not induce significant systematic error in data analysis for current detector sensitivity at the population level, while some special events still need further investigation.

4.2.3 Waveform difference in different parameter regions

In Sec 4.2.1, from the angle of data analysis, we discussed SNR's impact on waveform accuracy. What is more physically interesting is how the waveform accuracy varies with the intrinsic properties of the GW source. It is plausible that model accuracy decreases when the system includes some complex processes, such as a highly asymmetric mass ratio, high spin effects, high eccentricity and so forth. Accuracy may also drop when the contributions from higher modes increase, which usually happens to edge-on binaries [412, 413, 414].

We plot posterior samples of selected events and highlight the samples whose waveform difference is greater than 2 in Fig. 4.4. In the nearly equal mass region and small

spin region, `IMRPhenomXPHM` and `SEOBNRv4PHM` agree with each other and have waveform differences less than 2. However, when the mass ratio deviates from 1, or when spin parameters deviate from 0, the waveform difference grows and the waveform pair fails to pass the accuracy standard. For extrinsic parameters, we find that the waveform difference is largest when inclination angle $\theta_{\rm JN}$ for precessing systems is close to $\pi/2$. We attribute this to two causes: the contribution of higher modes increases when the source is edge-on, and the amplitude modulations caused by precession become increasingly visible, magnifying differences in the way precession is modelled [288, 412, 413, 414, 415, 416]. This is the reason why events like GW191109_010717 have a small fraction of posterior samples that pass the accuracy standard: estimations of their parameters mainly lead to low mass ratio, high spin regions or edge-on regions.

We then perform simulations of BBH events on the design sensitivity of Advanced LIGO [417]. We set the primary mass at $30M_\odot$, and mass ratio at $1, 0.8, 0.5$ and 0.2. The spin of each component is randomly generated: spin magnitude is uniformly distributed between 0 and 1, and the spin direction is isotropic. Inclination angle is isotropic as well. We neglect detector response functions and only include plus polarization here, which will not change our qualitative conclusions. For each mass ratio we simulate 6000 BBH events and calculate the waveform difference between `IMRPhenomXPHM` and `SEOBNRv4PHM`. Since waveform difference $\|\Delta\|$ is proportional to SNR, we introduce an SNR threshold above which $\|\Delta\|$ will be greater than 2. In Fig. 4.5, we plot the distributions of simulation parameters in the style of corner plot for different mass ratios, and the corresponding SNR thresholds in colors. We find the SNR threshold can reach 30 in the low spin and face-on region, but gradually drops below 10 as the spin parameters increase or $\theta_{\rm JN}$ tends to $\pi/2$. The change in mass ratio has the same effect, $\|\Delta\|$ can reach 2 at a smaller spin if the mass ratio is low. However, we find the $q = 0.2$ simulations can achieve a high SNR threshold for low-spin face-on sources, while high-spin or edge-on simulations are more likely to produce low SNR thresholds regardless of the mass ratio. Therefore, for current waveform modeling, spin effects and higher modes may need more improvements than low mass ratio cases. This coincides with our calculation on the asymmetric mass ratio but small spin event GW191219_163120. The disagreement in high-spin CBC waveforms and its impact on parameter estimation is also reported in Ref. [410].

Our simulation is consistent with the calculation for real events. Given GW events with SNRs ranging from 8 to 30 (for current detector sensitivity), those generated by nearly equal mass systems or low spin systems would have more $\|\Delta'_{\rm net}\| < 2$ samples, while the other events' posterior samples mostly fail our test, like GW191109_010717. Using the same method in Sec. 4.2.1 and comparing the current SNR threshold with the expected SNR of 3rd generation GW detectors, we can also conclude that, in general, the waveform accuracy should be improved for 3 to 4 orders of magnitude. However, for high spin and

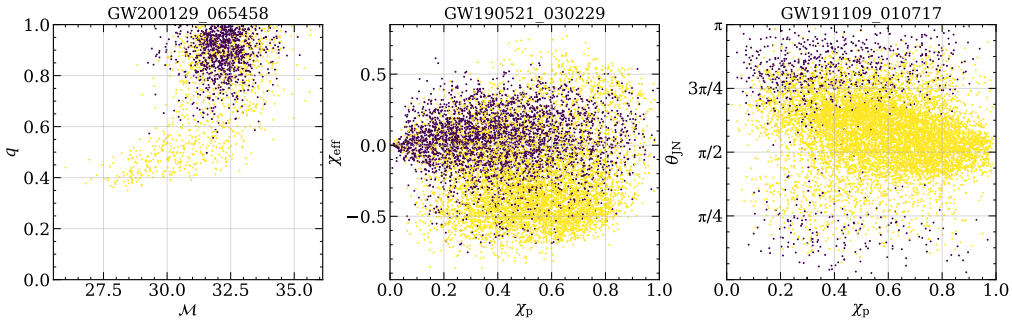


Figure 4.4: Posterior scatter of selected events. Yellow points represents samples with $\|\Delta'_{\rm net}\| > 2$, purple points are samples with $\|\Delta'_{\rm net}\| < 2$. We show three representative events with in two-dimension parameter plane $(\mathcal{M}, q)$, $(\chi_{\rm p}, \chi_{\rm eff})$ and $(\chi_{\rm p}, \theta_{\rm JN})$, respectively. It shows the inaccuracies mainly come from the high spin region, low mass ratio region, and edge-on region.

low mass ratio regions, as well as higher modes, the current waveform models may need more improvements. To calibrate waveform models, these regions might be where NR simulations are most needed for future waveform modeling.

4.3 Applications on BNS and NSBH waveforms

4.3.1 Real events

NSBH and BNS events are much less frequent than BBH events - only three events are generally considered as NSBH candidates in GWTC-2.1 and GWTC-3: GW191219_163120, GW200105_162426, and GW200115_042309, and two are considered as BNS events: GW170817 and GW190425_081805. Due to the complexity of these systems (e.g., highly asymmetric mass ratio, eccentricity for NSBH binaries, and matter effects for both types), some physical effects are yet to be included in their waveform models. Current NSBH waveform models of `IMRPhenom` and `SEOBNR` familes, `IMRPhenomNSBH` and `SEOBNRv4_ROM-NRTidalv2_NSBH` [418], are calibrated by non-spinning neutron star simulations and only allow aligned spins. For current BNS models, `IMRPhenomPv2_NRTidalv2` supports precessing spins while `SEOBNRv4T_surrogate` only supports aligned spins. Recent works have made `TEOBResumS` able to generate waveforms for precessing BNS systems with higher modes [419] , as well as waveforms for eccentric BNS systems [420], but they have not been applied to the GWTC-2.1 and GWTC-3.

For the three NSBH events, we calculate the $\|\Delta'_{\rm net}\|$ of their posterior samples generated by `IMRPhenomNSBH` and `SEOBNRv4_ROM_NRTidalv2_NSBH`. The fraction of $\|\Delta'_{\rm net}\| < 2$ samples are $99.4\%, 99.6\%$ and 100% for GW191219_163120, GW200105_162426, and GW200115_042309, respectively. Low SNR of these three events may contribute to the small waveform differences, but compared with the BBH events, lacking of precession

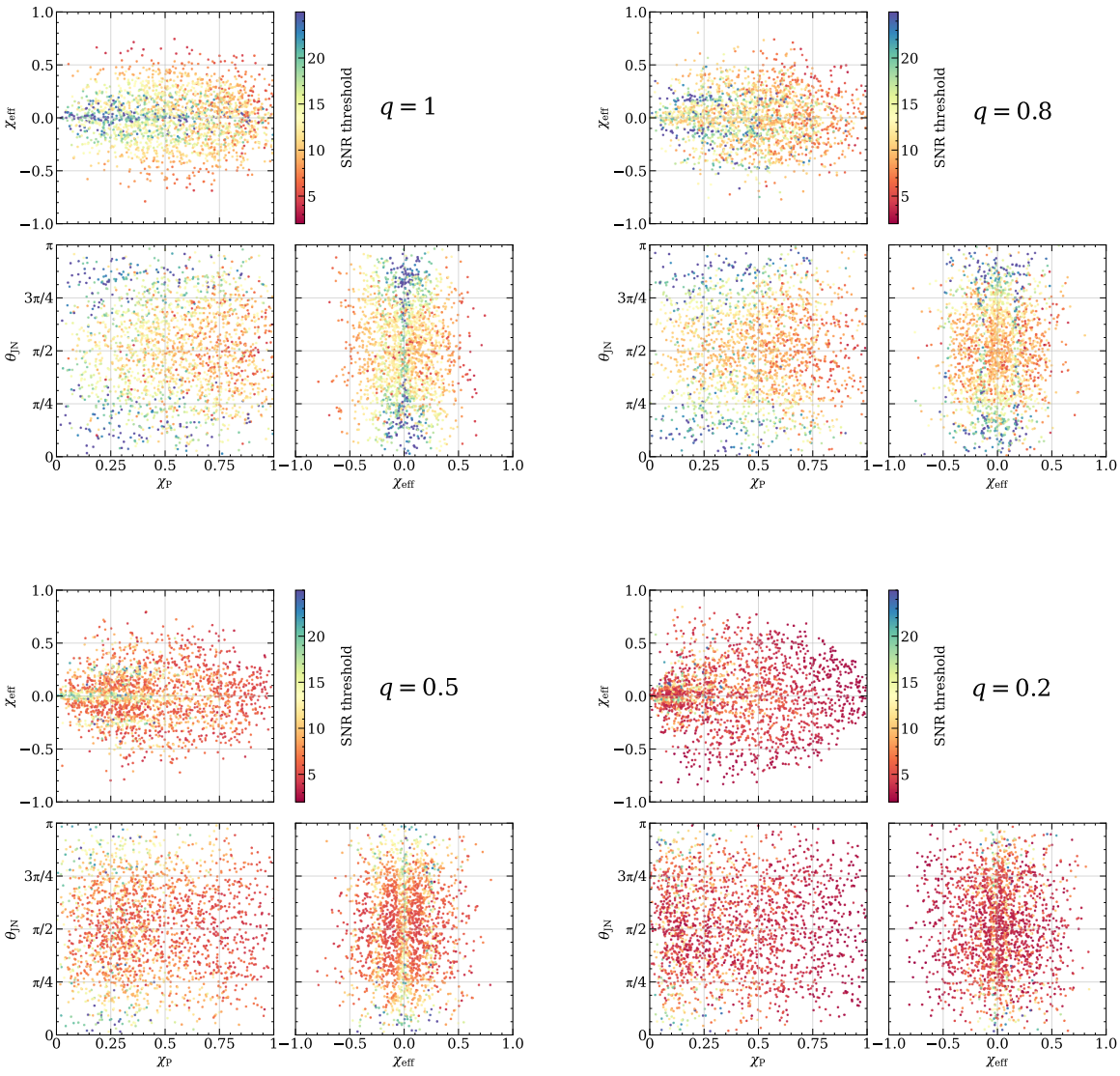


Figure 4.5: Simulations of BBH waveforms with random spin and inclination under LIGO design sensitivity. The primary mass is fixed at $30M_{\odot}$ and the mass ratio varies from 1 to 0.2, as shown in the top right corner of each figure. We calculate waveform difference $\|\Delta\|$ between `IMRPhenomXPHM` and `SEOBNRv4PHM` for each simulation and the SNR when $\|\Delta\|$ reaches 2. The SNR threshold is shown in different colors. Face-on events with smaller spins and equal masses tend to have a higher SNR threshold.

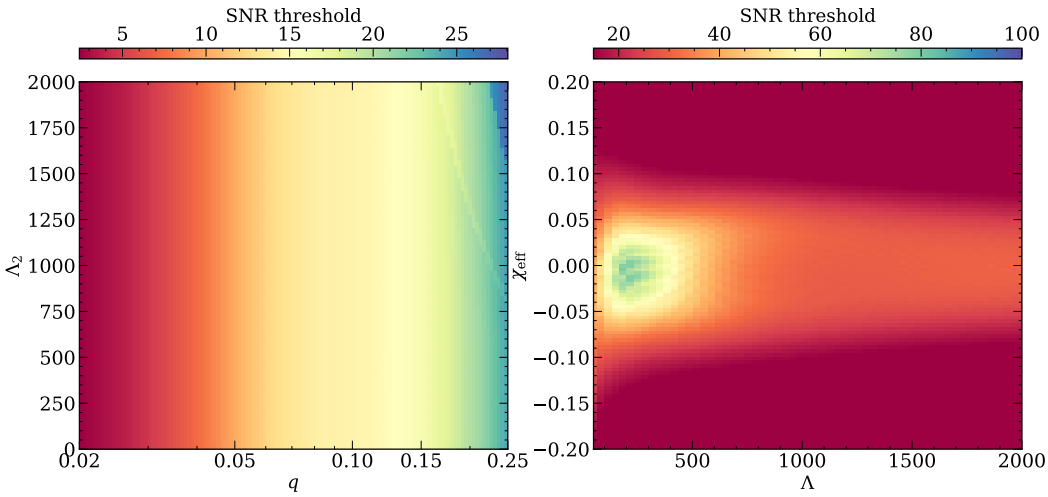


Figure 4.6: Left panel: Simulations of NSBH binaries in the mass ratio q - tidal deformability Λ_2 plane. The mass of the neutron star is fixed at 1.4 $M_\odot$, and we assume both components have zero spin. The colors in the plane represent the SNR threshold for the waveform difference between `IMRPhenomNSBH` and `SEOBNRv4_ROM_NRTidalv2_NSBH`, defined in the same way as before.
Right panel: Simulations of BNS binaries in the tidal deformability Λ - effective spin $\chi_{\rm eff}$ plane. We assume both neutron stars are 1.4 $M_\odot$ and they have the same spin and tidal deformability parameter. Colors represent SNR threshold for `IMRPhenomPv2_NRTidalv2` and `SEOBNRv4T_surrogate`.

effects and higher modes should be the decisive factors that make the waveform pair coincide, and it does not necessarily mean these models can describe general NSBH systems with high accuracy. As for BNS events, `IMRPhenomPv2_NRTidalv2` is the only model used in GWTC-2.1 and GWTC-3 that includes precession effects, it is not feasible to compare waveform difference of its posterior samples with others. Hence we do not include calculation of BNS waveforms for real events in this work.

4.3.2 Simulations

We perform simulations for NSBH and BNS systems respectively. For NSBH waveform models, we assume zero spin and secondary mass of 1.4 $M_\odot$. We change mass ratio between 0.02 and 0.25, and tidal deformability parameter between 0 and 2000. For BNS, we assume the two neutron stars are exactly the same: same mass 1.4 $M_\odot$, same tidal deformability parameter and spin. Then we change spin magnitude between -0.2 and 0.2, and tidal deformability parameter between 0 and 2000. We assume zero inclination for both systems. The results are shown in Fig. 4.6.

We find the main disagreement for NSBH waveform models lies in mass ratio, as

Fig. 4.6 shows the waveform difference drops with q but is insensitive to the tidal deformability parameter Λ_2 of the neutron star. The latter is because both approximants use the `NRTidalv2` [421] phase description to model the matter effects. SNR threshold can drop below 5 when q is small, but all the three NSBH candidates have SNRs lower than the thresholds indicated in the corresponding regions in Fig. 4.6. Note we assume zero spin in this simulation, but non-zero spin samples exist in the three NSBH candidates and would make extra contributions to waveform difference. Therefore, they still have a small fraction of $\|\Delta'_{\rm net}\| > 2$ samples. Given the SNR threshold in this simulation, NSBH waveform model accuracies (in terms of the mismatch from real waveform) also need an improvement of 3-4 orders of magnitude for future detection, leaving aside the unincluded physical effects.

As for BNS waveforms `IMRPhenomPv2_NRTidalv2` and `SEOBNRv4T_surrogate`, we change values of Λ and spin magnitude. We assume both components have the same aligned spin and mass, so the individual spin magnitude is equal to the effective spin. We find two waveform models agree with each other quite well in the $\Lambda < 500$, $\chi_{\rm eff} < 0.05$ region, with SNR thresholds up to 100. This is the region that coincides with our current understanding of neutron stars. However, when spin increases, the SNR threshold can drop below 20. This also implies accuracy of future waveform models should be improved by several orders of magnitude.

It should be noted that we might be underestimating the SNR threshold by not normalizing over the amplitude of waveform models. A slight constant amplitude difference in waveform models can lead to a different measurement of distance, but does not have large impacts on the intrinsic parameters. An alternative way of assessing waveform difference is mismatch, which is normalized by the waveform amplitudes. However, since the amplitude difference between waveform models is not significant, our conclusion of future accuracy requirements are consistent with previous works using mismatch [382].

We do not discuss further about NSBH and BNS waveform models, for we suppose the number of real events is not enough for us to perform analysis on the population level, and further work on precession, higher modes, and even eccentricity should be done for more NSBH and BNS waveform models.

4.4 Summary and discussions

In this chapter, we developed a diagnostic test for the presence of waveform mismodeling. This extends the work of Ref. [405] to realistic analyses. While Ref. [405] suggests a waveform model should have an error (as a vector) shorter than 1 to be accurate enough, we introduce two waveform models and find their difference should be shorter than 2 if they are both accurate enough. This method frees accuracy evaluation from the unknown

true waveform, and it enables the evaluation to be performed in larger, continuous regions in the parameter space: the regions where waveform models can work, rather than where NR simulations are done. We should note that our method can only tell the existence of inaccurate waveform models. It can not tell which one (or both) is (are) inaccurate if the pair fails, or guarantee any accuracy when the pair do not fail. The key idea is: If two models have significant differences, they can not be both accurate enough, but when the difference is small, we can not rule out the possibility that two models are making similar mistakes.

For BBH waveform models, we choose the state-of-the-art models from `IMRPhenom` and `SEOBNR` family, `IMRPhenomXPHM` and `SEOBNRv4PHM` for illustration. We have applied our test to existing parameter estimates from the GWTC-3 and GWTC-2.1 (which used the waveform models mentioned above), and found differences in the results of data analysis from different waveform models. The samples that fail our test are mostly located in the low mass ratio, high spin or egde-on regions in the parameter space, which means waveform models become less reliable in these regions. Our simulations agree with this: the waveform difference between `IMRPhenomXPHM` and `SEOBNRv4PHM` can reach the threshold 2 when SNR is less than 10 in high spin regions and edge-on regions; waveform difference increases in low mass ratio region the as well. We also note that spin effects and inclinations (higher modes) are more problematic for waveform modeling than mass ratio. This points out where NR simulations are needed most for future waveform calibration.

We have investigated the correlation between the waveform difference and inconsistency of parameter estimation samples given by different waveform models. The latter is measured by the J-S divergence. For the GWTC-3 and GWTC-2.1 posterior samples, we find when the fraction of $\Delta < 2$ samples is less than 40%, it is more likely to obtain a J-S divergence larger than most other events, which is a sign of underlying systematic errors caused by waveform error. We also note that the inverse is not necessarily true, as the waveform model is not the only factor that can influence the generation of posterior samples, but nonetheless it is always helpful to have one of the factors checked. Since multi-waveform analysis is becoming a standard way of reducing systematic errors in parameter estimation of GW sources, we suggest that waveform difference analysis can be used as a real-time quantitative check in the parameter estimation workflow.

For NSBH waveforms, we select `IMRPhenomNSBH` and `SEOBNRv4_ROM_NRTidalv2_NSBH`, the two models used in GWTC-3 and GWTC-2.1 parameter estimation. The posterior samples of the 3 NSBH candidates have small NSBH waveform differences compared to BBH waveforms. We credit this to the fact that these waveform models do not include non-aligned spins or higher modes as BBH waveforms. As expected, we find waveform difference increases when the mass ratio decreases in our simulation. The SNR threshold drops below 10 when the mass ratio is less than 0.05, indicating that more calibrations

are needed for this region, leaving aside the lack of some other physical effects.

For BNS waveforms, we have not applied our test on real events samples, for only `IMRPhenomPv2_NRTidalv2` is used in GWTC-2.1 and -3 data analysis, and we could not find another comparable model to be paired with it. We simulate aligned spin BNSs for `IMRPhenomPv2_NRTidalv2` and `SEOBNRv4T_surrogate` instead. We find the systematic differences between the approximants we examined are small in the region where $\Lambda < 500$, and $|\chi_{\rm eff}| < 0.05$, which should be the case for our current understanding of neutron stars. However, there are some differences when $\Lambda < 50$ where the waveforms appear to diverge again. In the high spin regions, the SNR threshold drops below 20, which can not meet future high SNR detections.

The waveform difference is related to the widely used overlap (or mismatch) through Eqs. 4.15 and 4.16. If we assume two models are well-calibrated and have comparable errors, we can give a rough estimate of future waveform accuracy requirements. This complements previous works on waveform accuracy [382]. Looking at the SNR thresholds for the three types of waveforms, we know the current waveform accuracy is not enough for future high SNR detections where SNR can reach up to 1000. The mismatch from the real waveform needed to be reduced by at least 3 orders of magnitude. This is consistent with previous work.

This method can be extended to more complex GW waveform models for future GW detection, such as waveforms including eccentricity. We can perform this analysis as long as there are at least two waveform models with similar accuracy and which include the same physical parameters. Our method can work beyond the NR calibration range, thus it can be an efficient way to study the waveform models' extrapolation performance. This may also be a guide to where NR simulations are most needed in the parameter space.

Finally, this waveform accuracy assessment method has been coded up in the package `WaveCheck` [1] and applied on O4 online parameter estimations. `WaveCheck` regularly downloads the latest parameter estimation result files and analyzes the waveform difference between the start-of-the-art waveforms (which, in O4, are `IMRPhenomXPHM`, `SEOBNRv5PHM`, and `NRSur7dq4`), and generates a report page[2] including Δ, mismatch, and figures similar to Fig. 4.4. The results are not public at the time of writing this thesis since O4 is still ongoing, but they will eventually be.

[1] https://git.ligo.org/qian.hu/wavecheck_o4
[2] https://ligo.gravity.cf.ac.uk/~qian.hu/wavecheck/

Chapter 5

Rapid pre-merger localization of BNS in the 3G era

Pre-merger localization of Binary Neutron Star (BNS) is one of the most important scientific goals for the Third-Generation (3G) Gravitational Wave (GW) detectors. It will enable the electromagnetic observation of the whole process of BNS coalescence, especially for the pre-merger and merger phases which have not been observed yet, opening a window for a deeper understanding of compact objects. To reach this goal, we describe a novel combination of multi-band matched filtering and semi-analytical localization algorithms to achieve early-warning localization of long BNS signals in 3G detectors. Using our method we are able to efficiently simulate one month of observations with a three-detector 3G network, and show that it is possible to provide accurate sky localizations more than 30 minutes before the merger. Our simulation shows that there could be ~ 10 (~ 100) BNS events localized within 100 deg^2, 20 (6) minutes before the merger, per month of observation.

We will discuss the necessity of fast localization in Sec. 5.1, and introduce the fast localization algorithm SealGW in Sec. 5.2. In Sec. 5.3, we will build a multi-banding scheme and extend the localization algorithm to pre-merger early warning for long GW signals. Tests and diagnoses, including localization statistics, will be given in Sec. 5.4. We summarize the results in Sec. 5.5.

5.1 The importance of fast localization

Since the first direct Gravitational Wave (GW) detection of the coalescence of Binary Neutron Star (BNS) GW170817 [116] and its Electromagnetic (EM) counterparts [49], multi-messenger observation of coalescing compact binaries has become an important tool for astrophysics. Joint GW-EM observations can provide a comprehensive understanding of the formation and evolution of BNS, and shed light on physics around compact objects [135, 136, 422, 423, 424, 425, 426, 427, 428, 429, 430].

Rapid GW detection and accurate localization is key to joint GW-EM observations, as most EM facilities need direction from the GW observation. In addition to capturing the afterglow of BNS coalescences, early EM observations could offer unique insights into phenomena in BNS that happen prior to or near the merger (e.g. tidal disruptions, magnetosphere interactions, r-process nucleosynthesis) and help build a picture of the entire evolution of kilonova in multiple frequency bands [48, 425, 431, 432, 433, 434]. Early detection and localization of BNS are therefore of great importance in GW astronomy. It has previously been demonstrated that there is a non-zero probability of detecting and localizing pre-merger BNS events with current [435, 436] and near-future [437, 438, 439, 440] GW observatories, typically within seconds to one minute before merger. [301] has recently investigated early warning for the fourth observing run of LIGO-Virgo-KAGRA collaboration (LVK) collaboration, with multiple detection pipelines already equipped for early warning searches [300, 441, 442, 443]. Machine learning based methods for early warning detection are also making rapid progresses [444, 445, 446, 447]

Being limited by low sensitivities below 20 Hz, the BNS signal is only detectable for few minutes in current GW detectors. Given the communication time delay between multi-messenger community and the $\sim 10-100$ seconds slew time of modern telescopes [440], it is basically impossible to capture pre-merger or near-merger transients from BNS coalescence without fore-warning. Several Third-Generation (3G) GW detectors have been proposed, including Einstein Telescope (ET) [166] and Cosmic Explorer (CE) [168, 169], with low-frequency sensitivities significantly improved. These would allow us to detect BNS signals more than 30 minutes before the merger, rendering precise early warning localization possible [167, 171, 448, 449, 450, 451].

However, data analysis of BNS in 3G detectors can be challenging. The long signal makes matched filtering extremely expensive to perform, and it is modulated by changes in the antenna response functions due to the Earth's rotation. Besides, the large number of signals (see, e.g., Tab. 3.1) requires an efficient localization algorithm to rapidly localize the source with minimal computational resources. In this chapter, we will present a multi-banding scheme and an efficient localization algorithm SealGW to achieve the pre-merger localization for the 3G detectors.

5.2 Semi-analytical localization algorithm for GWs

In this section, we introduce the localization algorithm Semi-Analytical localization for Gravitational Waves (SealGW) [4]. Fast localization aims to obtain the posterior distribution $p(\alpha, \delta|\boldsymbol{d})$, where α and δ are the right ascension and declination, respectively. The sky location posterior is actually the marginalized posterior

$$p(\alpha, \delta|\boldsymbol{d}) = \int p(\boldsymbol{\theta}) p(\boldsymbol{d}|\boldsymbol{\theta}) d\boldsymbol{\theta}_{\mathbf{nui}}, \tag{5.1}$$

where $\boldsymbol{\theta}_{\mathbf{nui}}$ are the nuisance parameters and $\boldsymbol{\theta} = \{\alpha, \delta, \boldsymbol{\theta}_{\mathbf{nui}}\}$. In principle, $\boldsymbol{\theta}_{\mathbf{nui}}$ should include intrinsic parameters, but intrinsic parameters are initially estimated by matched filtering searches and their errors are semi-independent with errors in sky localization [112, 452, 453, 454], so they are treated as perfectly known in online fast localization. Therefore, the nuisance parameters are the five remaining extrinsic parameters: $\boldsymbol{\theta}_{\mathbf{nui}} = \{d_L, \iota, \psi, \phi_c, t_c\}$.

`Bayestar` [452], the current standard fast localization algorithm used by LVK, performs a five-fold numerical marginalization over the nuisance extrinsic parameters to obtain the sky location. SealGW aims to do this "as analytically as possible" so that it can be faster and less demanding on hardware. I will briefly introduce how SealGW works in this section.

5.2.1 Rearranging likelihood and parameters

To perform analytical marginalization for Eq. 5.1, we should rearrange the form of the likelihood. Define *amplitude modulation functions* $G^{(i)}_{+,\times}$:

$$G^{(i)}_{+,\times}(\alpha, \delta, t_c) = F^{(i)}_{+,\times}(\alpha, \delta, \psi = 0, t_c), \tag{5.2}$$

it is related to the antenna response functions by [106]

$$\begin{aligned} F_+(t) &= \sin\zeta \left[G_+(t)\cos 2\psi + G_\times(t)\sin 2\psi\right] \\ F_\times(t) &= \sin\zeta \left[G_\times(t)\cos 2\psi - G_+(t)\sin 2\psi\right], \end{aligned} \tag{5.3}$$

where ζ is the angle between two arms of the interferometer. This relation is from the fact that the polarization angle ψ simply rotates the wave frame. With the waveform decomposition Eq. 2.16, a detector's response to GWs Eq. 1.110 can be written as

$$h^{(i)} = (G^{(i)}_+, G^{(i)}_\times)\mathbf{A_c} h_c + (G^{(i)}_+, G^{(i)}_\times)\mathbf{A_s} h_s, \tag{5.4}$$

where $^{(i)}$ denotes the i'th detector and

$$\begin{aligned}\mathbf{A} &= \begin{pmatrix}\mathbf{A_c} & \mathbf{A_s}\end{pmatrix} = \begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix} \\ &= \frac{1\text{Mpc}}{r}\begin{pmatrix}\cos 2\psi & \sin 2\psi\\ -\sin 2\psi & \cos 2\psi\end{pmatrix}\begin{pmatrix}\frac{1+\cos^2\iota}{2} & \\ & \cos\iota\end{pmatrix}\begin{pmatrix}\cos\phi_c & \sin\phi_c\\ -\sin\phi_c & \cos\phi_c.\end{pmatrix}\end{aligned} \tag{5.5}$$

The likelihood function for one detector can be written as:

$$\begin{aligned}p(d^{(i)} \mid \boldsymbol{\theta}) &\propto e^{-(d^{(i)}-h^{(i)}|d^{(i)}-h^{(i)})/2} \\ &\propto e^{(d^{(i)}|h^{(i)})-\frac{1}{2}(h^{(i)}|h^{(i)})},\end{aligned} \tag{5.6}$$

and for a network,

$$p(\mathbf{d} \mid \boldsymbol{\theta}) \propto e^{(\mathbf{d^T}|\mathbf{h})-\frac{1}{2}(\mathbf{h^T}|\mathbf{h})} \tag{5.7}$$

where

$$\mathbf{d} = \begin{pmatrix}d^{(1)}(t+\tau^{(1)})\\ d^{(2)}(t+\tau^{(2)})\\ \vdots \\ d^{(N)}(t+\tau^{(N)})\end{pmatrix}, \quad \mathbf{h} = \begin{pmatrix}h^{(1)}(t+\tau^{(1)})\\ h^{(2)}(t+\tau^{(2)})\\ \vdots \\ h^{(N)}(t+\tau^{(N)})\end{pmatrix}, \tag{5.8}$$

where $\tau^{(i)}$ is to take into account different arrival times of the signal. Extending Eq. 5.7 with Eq. 5.4 and split the matrix $\mathbf{A}$ into $\mathbf{A_c} = (A_{11}, A_{21})^{\mathrm{T}}$ and $\mathbf{A_s} = (A_{12}, A_{22})^{\mathrm{T}}$, we have the network likelihood function:

$$p(\mathbf{d} \mid \boldsymbol{\theta}) \propto \prod_{x=\{c,s\}} e^{\left(\mathbf{d^T}|\mathbf{G}\mathbf{A_x}h_x\right)-\frac{1}{2}\left(\mathbf{A_x^T}\mathbf{G^T}h_x|\mathbf{G}\mathbf{A_x}h_x\right)}. \tag{5.9}$$

Define σ as the template norm:

$$\begin{aligned}(h_c|h_s) &= 0 \\ \sigma^{(i)} \equiv \sqrt{(h_c|h_c)}\,|_{r=1\text{Mpc}} &= \sqrt{(h_s|h_s)}\,|_{r=1\text{Mpc}}\end{aligned} \tag{5.10}$$

then the likelihood becomes

$$p(\mathbf{d} \mid \boldsymbol{\theta}) \propto \prod_{x=\{c,s\}} e^{\left(\mathbf{d^T}|\mathbf{H_x}\right)\mathbf{G}_\sigma\mathbf{A_x}-\frac{1}{2}\mathbf{A_x^T}\mathbf{G_\sigma^T}\mathbf{G}_\sigma\mathbf{A_x}}. \tag{5.11}$$

where $\mathbf{G}_\sigma$ is

$$\mathbf{G}_\sigma = \begin{pmatrix}G_+^{(1)}\sigma^{(1)} & G_\times^{(1)}\sigma^{(1)}\\ G_+^{(2)}\sigma^{(2)} & G_\times^{(2)}\sigma^{(2)}\\ \vdots & \vdots \\ G_+^{(N)}\sigma^{(N)} & G_\times^{(N)}\sigma^{(N)}\end{pmatrix}. \tag{5.12}$$

and $\mathbf{H_{c,s}}$ is the normalized template given as:

$$\mathbf{H_{c,s}} = \mathrm{diag}\left(\frac{h_{c,s}}{\sigma^{(1)}}, \frac{h_{c,s}}{\sigma^{(2)}}, \ldots, \frac{h_{c,s}}{\sigma^{(N)}}\right), \tag{5.13}$$

where *diag* denotes diagonal matrix. $\mathbf{H_{c,s}}$ is also used to compute the matched filtering SNR, as we have derived in Eq. 2.26

$$\boldsymbol{\rho} = (\mathbf{H_c} \mid \mathbf{d}) + i(\mathbf{H_s} \mid \mathbf{d}). \tag{5.14}$$

The i'th element of $\boldsymbol{\rho}$ is the SNR time series of the i'th detector. It is a complex time series and we use its modulus as the SNR. The network SNR is defined as $\sqrt{\boldsymbol{\rho}^{\boldsymbol{T}}\boldsymbol{\rho}^{*}}$.

Eq. 5.11 tells us the likelihood can be computed by the match filtering SNR, and can be written as a Gaussian function of the matrix $\mathbf{A}$. The former helps the fast online localization as the SNR timeseries is available in real-time and it contains an initial estimate of intrinsic parameters which we have ignored in marginalization. In other words, the likelihood computed with SNR timeseries is conditioned by the template's intrinsic parameters. The Gaussianity encourages us to use the following parameter conversion since $\mathbf{A}$ is dependent on extrinsic parameters $\{d_L, \iota, \psi, \phi_c\}$:

$$(t_c, \alpha, \delta, \iota, \phi_c, d_L, \psi) \rightarrow (t_c, \alpha, \sin\delta, A_{11}, A_{21}, A_{12}, A_{22}). \tag{5.15}$$

In fact, a number of works have employed the transformation to $\mathbf{A}$, especially in the area of pulsar searches [455, 456, 457]. From Eq. 5.5, we have

$$\begin{aligned}
A_{11} &= \frac{1}{d_L}\left(\frac{1+\cos^2\iota}{2}\cos 2\psi\cos\phi_c - \cos\iota\sin 2\psi\sin\phi_c\right),\\
A_{21} &= -\frac{1}{d_L}\left(\frac{1+\cos^2\iota}{2}\sin 2\psi\cos\phi_c + \cos\iota\cos 2\psi\sin\phi_c\right),\\
A_{12} &= \frac{1}{d_L}\left(\frac{1+\cos^2\iota}{2}\cos 2\psi\sin\phi_c + \cos\iota\sin 2\psi\cos\phi_c\right),\\
A_{22} &= -\frac{1}{d_L}\left(\frac{1+\cos^2\iota}{2}\sin 2\psi\sin\phi_c - \cos\iota\sin 2\psi\cos\phi_c\right).
\end{aligned} \tag{5.16}$$

Now we need to marginalize over four elements of $\mathbf{A}$ and time t_c. Thanks to the Gaussianity, the integral over $\mathbf{A}$ can be performed analytically.

5.2.2 Prior setting

We should assign the prior distributions for $\mathbf{A}$, t_c, and (α, δ). The prior of t_c is assigned as uniform distribution $\pm$10ms around the trigger time, and (α, δ) is chosen to be isotropic,

i.e,

$$\begin{aligned} p(\alpha) &\propto 1, \\ p(\sin\delta) &\propto 1, \\ p(t_c) &\propto 1, \end{aligned} \tag{5.17}$$

which means they can be treated as constants in the posterior probability density.

The prior distribution of A_{ij} is determined by Monte Carlo simulations. For example, we simulate 50000 events in an LHV detector network with the O2 sensitivity, and calculate A_{ij}. The result is shown in Fig. 5.1. We find the following characteristics of the distribution of A_{ij}:

- Each A_{ij} follows the similar distribution.
- When SNR is high (> 8), A_{ij} follows bimodal distribution, and the location of the peak depends on SNR.
- The diagonal and off-diagonal elements, (A_{11}, A_{22}) and (A_{21}, A_{12}), are correlated, while other elements are entirely uncorrelated with each other.

Based on the first point, we assume all A_{ij}s follow the same distribution. Since a GW signal with network SNR < 8 is usually not considered as a successful detection, we ignore the low-SNR cases and focus on the bimodal distribution of A_{ij}. We adopt a symmetric bimodal prior distribution for A_{ij} with a superposition of two Gaussian functions:

$$p(A_{ij}) \propto e^{-\frac{(A_{ij}-\mu)^2}{2\sigma^2}} + e^{-\frac{(A_{ij}+\mu)^2}{2\sigma^2}}, \tag{5.18}$$

where μ and σ will be derived numerically for a given astrophysical model of the extrinsic parameters for different ranges of SNR. For distributions of A_{ij} in each SNR bin with the length of 2, we use the least square method to obtain the best-fit μ and σ. Comparing the best-fit values in different SNR bins, we find μ and σ have a linear relation with the network SNR:

$$\begin{aligned} \mu &= 0.0004860\ \mathrm{SNR} - 0.0007827, \\ \sigma &= 0.0002733\ \mathrm{SNR} + 0.00005376, \end{aligned} \tag{5.19}$$

as shown in Fig. 5.2.

Fig. 5.3 shows the comparison between the distribution of A_{ij} in different SNR bins and the bimodal prior distribution calculated using Eq. 5.19 at the central value of the SNR bin. The bimodal distribution with empirical relation is sufficient to reconstruct A_{ij}s' distribution.

The bimodal distribution of A_{ij} originates from the selection effect of high SNR GW events. The amplitude of the GW increases monotonically with $|\cos\iota|$ and $1/d_L$, therefore, when we select GW events in a higher SNR range, sources with larger $|\cos\iota|$ values and

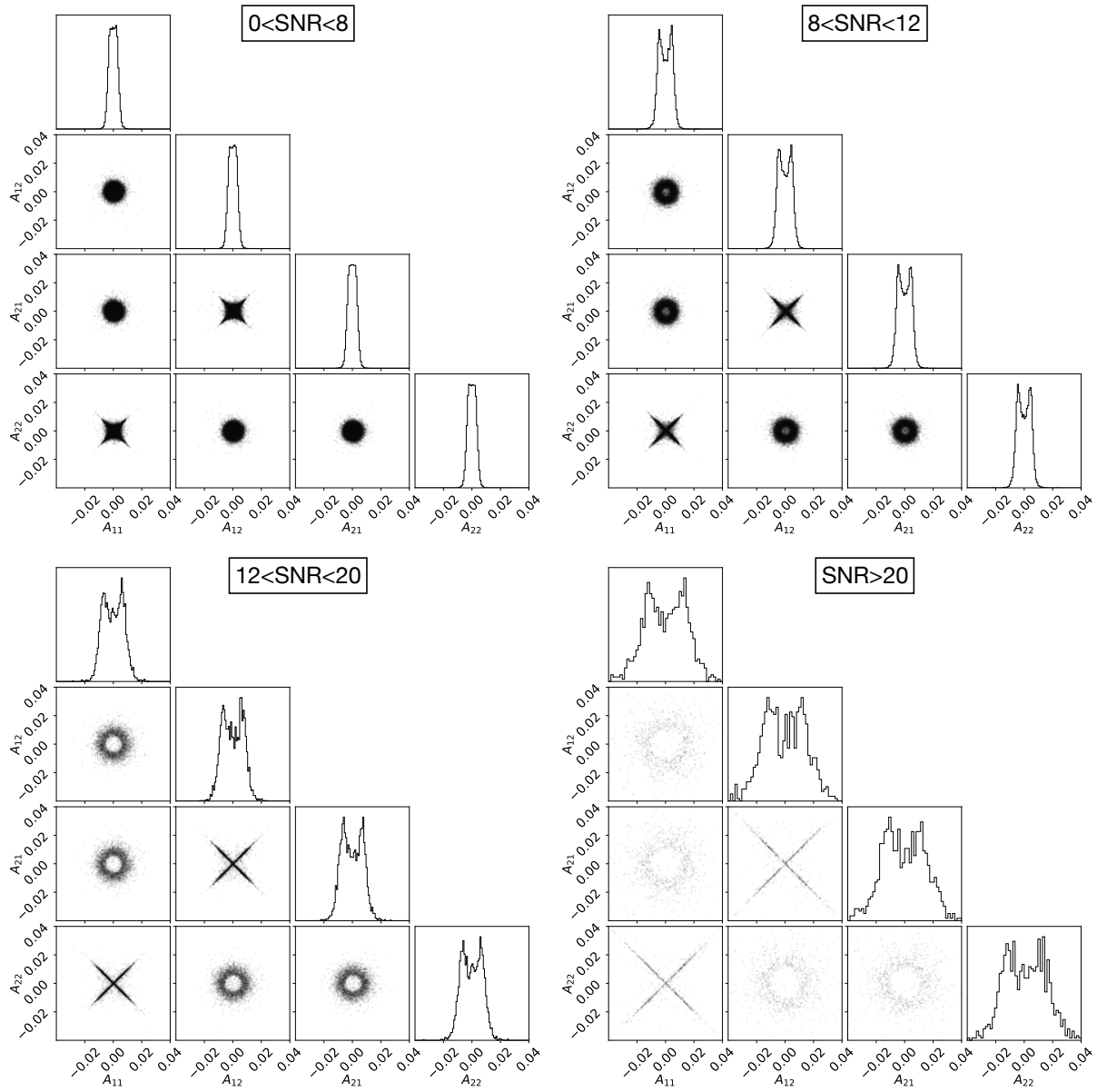


Figure 5.1: Corner plot of elements of **A** generated from 50000 simulations in O2 sensitivity. We categorized A_{ij} samples according to the HLV network SNR and plotted them in 4 subfigures.

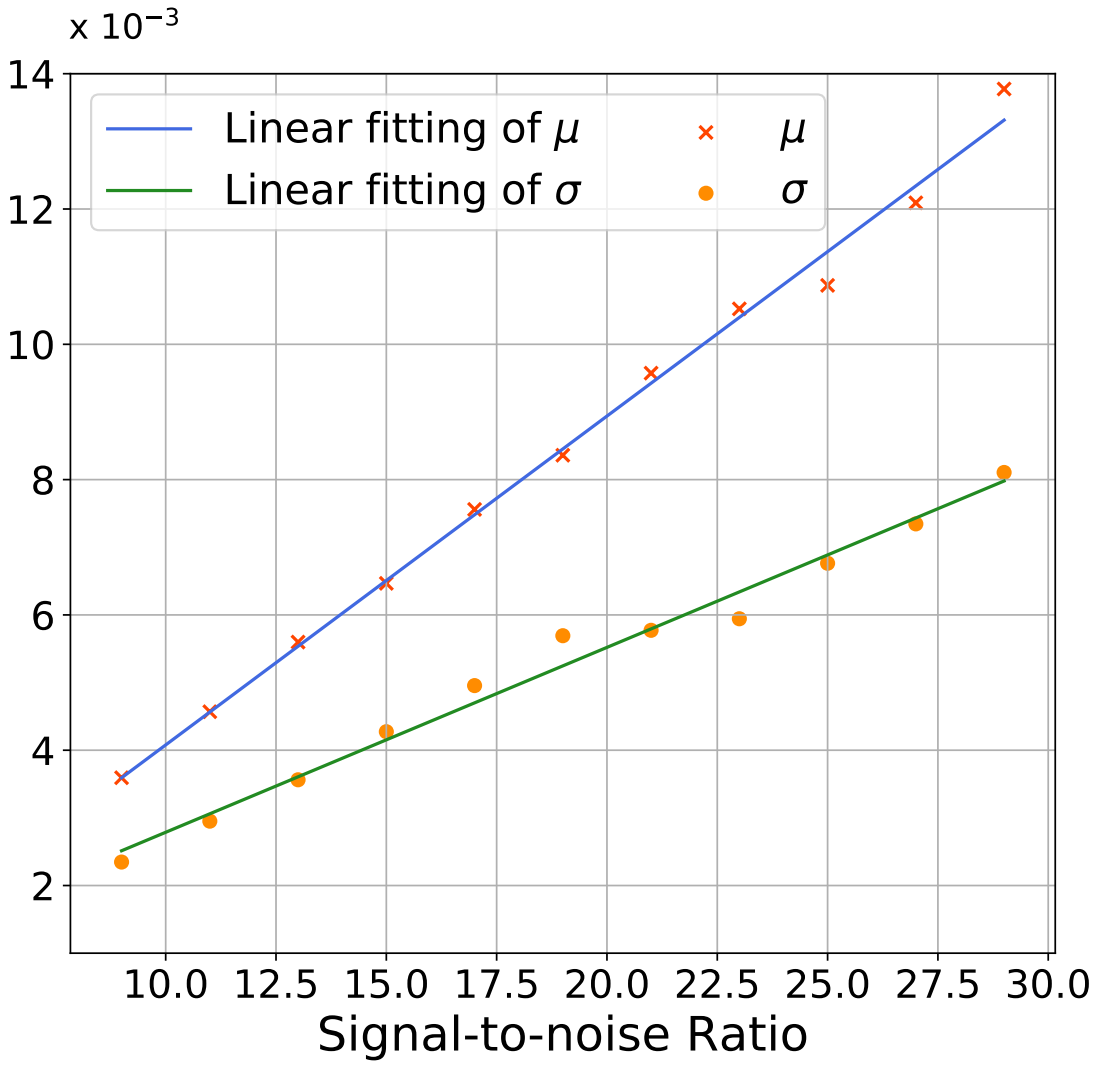


Figure 5.2: Linear relation between the best-fit μ, σ and SNR. Points of the best-fit values are plotted at the center of their corresponding SNR bin.

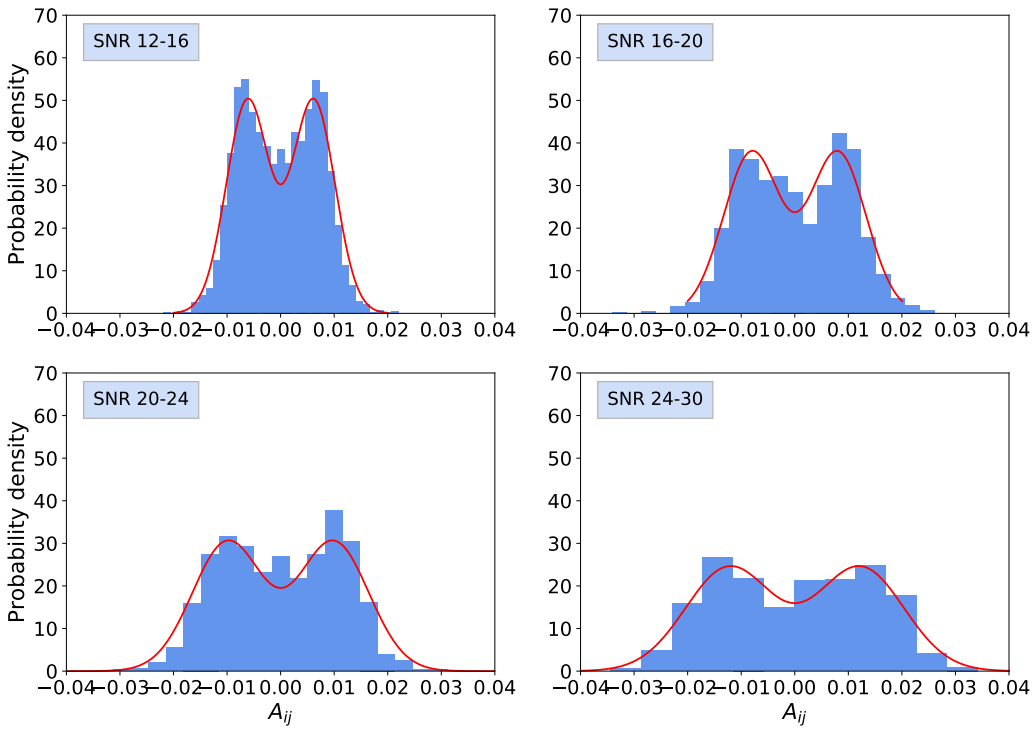


Figure 5.3: Comparison between A_{ij}'s simulated distribution and empirical approximation. Blue bars are A_{ij}'s distribution in our simulation; four subplots represent four different SNR ranges. Red lines are empirical distributions with SNRs equal to the center value of SNR bins.

smaller d_L are more likely to be selected. This also implies that $\cos\iota$ tends to have a symmetric bimodal distribution [106], which results in the bimodal distribution of A_{ij} (Eq. 5.16). Physically, the bimodality of $\cos\iota$ originates from the opposite handedness of the binary orbit, which corresponds to two different inclination angles: ι and $\pi - \iota$. On the other hand, small distance d_L corresponds to larger values of A_{ij}, hence the peaks in A_{ij}'s distribution move outside (towards $+x$ and $-x$ directions from the center). This is why A_{ij} follows bimodal distribution and is dependent on SNR.

The correlation between A_{ij}s can be explained as follows. As discussed above, for high SNR events we have $|\cos\iota| \to 1$. According to Eq. 5.16, when $\cos\iota \to 1$, $A_{11} \to A_{22}$, $A_{12} \to -A_{21}$; when $\cos\iota \to -1$, $A_{11} \to -A_{22}$, $A_{12} \to A_{21}$. Plus $\cos\iota$ has the same probability to be 1 or -1, as a result, two symmetric crosses are shown in each corner plot of Fig. 5.1.

As most A_{ij} samples in Fig. 5.1 are positioned at the diagonal cross, we further assume that A_{22} has half a chance to be A_{11} and another half to be $-A_{11}$ for detectable GW events whose SNRs are usually high, i.e.,

$$p(A_{22}|A_{11}) = \frac{\delta(A_{22} - A_{11}) + \delta(A_{22} + A_{11})}{2}, \tag{5.20}$$

where $\delta()$ is the δ function: $\delta(0) = 1$ and $\delta(x) = 0$ otherwise. For the same reason,

$$p(A_{12}|A_{21}) = \frac{\delta(A_{21} - A_{12}) + \delta(A_{21} + A_{12})}{2}. \tag{5.21}$$

This approximation is also adopted in previous work on fast GW source localization [458], in which the authors show the probability distribution of $\cos\iota$ of detectable GW events has strong peaks for $\cos\iota = \pm 1$. Note the correlation only exists in diagonal and off-diagonal elements of $\mathbf{A}$, we have

$$\begin{aligned} p(\mathbf{A}) &= p(A_{11}, A_{12}, A_{21}, A_{22}) \\ &= p(A_{11}, A_{22})p(A_{21}, A_{12}) \\ &= p(A_{11})p(A_{22}|A_{11})p(A_{21})p(A_{12}|A_{21}), \end{aligned} \tag{5.22}$$

where $p(A_{11})$ and $p(A_{21})$ are given in Eq. 5.18 while $p(A_{22}|A_{11})$ and $p(A_{12}|A_{21})$ are given in Eq. 5.20 and 5.21. However, we should point out that Eq. 5.20 is an approximation for detectors that are not very sensitive, like the O2 sensitivity. For the 3G detectors, the above arguments are not valid anymore and we assume A_{ij}s are independent for simplicity, i.e.,

$$\begin{aligned} p(\mathbf{A}) &= p(A_{11}, A_{12}, A_{21}, A_{22}) \\ &= p(A_{11})p(A_{22})p(A_{21})p(A_{12}). \end{aligned} \tag{5.23}$$

Since this chapter focuses on the application on 3G detectors, we will ignore the correlation

between A_{ij}s and go with the prior Eq. 5.23. The results and equations of correlated A_{ij} can be found in Ref. [4].

5.2.3 Semi-analytical marginalization

The posterior for sky location is

$$p(\alpha,\delta|\mathbf{d}) = \int dt_c \int d^4\mathbf{A} p(\alpha,\delta,t_c,\mathbf{A} \mid \mathbf{d}), \tag{5.24}$$

where $d^4\mathbf{A} = dA_{11}dA_{12}dA_{21}dA_{22}$. For simplicity, we define

$$\begin{cases} \mathbf{M} = \mathbf{G}_\sigma^{\mathrm{T}}\mathbf{G}_\sigma \\ \mathbf{J}_\mathbf{x}^{\mathrm{T}} = \left(\mathbf{d}^{\mathrm{T}} \mid \overline{\mathbf{H}}_\mathbf{x}\right)\mathbf{G}_\sigma \end{cases} \tag{5.25}$$

where $x = \{c, s\}$. $\mathbf{M}$ is a positive-definite 2×2 matrix, and $\mathbf{J}_\mathbf{x}$ is a column vector of length of two, with two components $J_{\mathrm{x}1}$, $J_{\mathrm{x}2}$. The posterior can be written as

$$\begin{aligned} p\left(\alpha,\delta \mid \mathbf{d}\right) &\propto \int_{t_c-T}^{t_c+T} dt_c \int d^4\mathbf{A} \exp\left\{\sum_{x=\{c,s\}} -\frac{1}{2}\mathbf{A}_\mathbf{x}^{\mathrm{T}}\mathbf{M}\mathbf{A}_\mathbf{x} + \mathbf{J}_\mathbf{x}^{\mathrm{T}}\mathbf{A}_\mathbf{x}\right\} \\ &\times \prod_{x=\{c,s\}} \left(e^{-\frac{(A_{x1}-\mu)^2}{2\sigma^2}} + e^{-\frac{(A_{x1}+\mu)^2}{2\sigma^2}}\right)\left(e^{-\frac{(A_{x2}-\mu)^2}{2\sigma^2}} + e^{-\frac{(A_{x2}+\mu)^2}{2\sigma^2}}\right) \\ &\propto \int_{t_c-T}^{t_c+T} dt_c \prod_{x=\{c,s\}} \int d\mathbf{A}_\mathbf{x} \sum_{i=1}^{4} \exp\left\{-\frac{1}{2}\mathbf{A}_\mathbf{x}^{\mathrm{T}}\left(\mathbf{M}+\frac{1}{\sigma^2}\mathbf{I}\right)\mathbf{A}_\mathbf{x} + \mathbf{J}^{(\mathbf{i})\mathrm{T}}{}_\mathbf{x}\mathbf{A}_\mathbf{x}\right\} \\ &\propto \int_{t_c-T}^{t_c+T} dt_c \frac{(2\pi)^2}{\det\left(\mathbf{M}+\frac{1}{\sigma^2}\mathbf{I}\right)} \prod_{x=\{c,s\}} \sum_{i=1}^{4} \exp\left\{\frac{1}{2}\mathbf{J}^{(\mathbf{i})\mathbf{T}}{}_\mathbf{x}\left(\mathbf{M}+\frac{1}{\sigma^2}\mathbf{I}\right)^{-1}\mathbf{J}^{(\mathbf{i})}{}_\mathbf{x}\right\}, \end{aligned} \tag{5.26}$$

where μ, σ are the parameters in the prior, $\mathbf{I}$ is the unitary matrix, and

$$\mathbf{J}_\mathbf{x}^{(\mathbf{i})} = \mathbf{J}_\mathbf{x} + \boldsymbol{\alpha}^{(\boldsymbol{i})},\ i = 1,2,3,4. \tag{5.27}$$

$\boldsymbol{\alpha}^{(\boldsymbol{k})}$is given by

$$\begin{aligned} \boldsymbol{\alpha}^{(\mathbf{1})} &= \begin{pmatrix} \mu/\sigma^2 \\ \mu/\sigma^2 \end{pmatrix}, \boldsymbol{\alpha}^{(\mathbf{2})} = \begin{pmatrix} \mu/\sigma^2 \\ -\mu/\sigma^2 \end{pmatrix}, \\ \boldsymbol{\alpha}^{(\mathbf{3})} &= \begin{pmatrix} -\mu/\sigma^2 \\ \mu/\sigma^2 \end{pmatrix}, \boldsymbol{\alpha}^{(\mathbf{4})} = \begin{pmatrix} -\mu/\sigma^2 \\ -\mu/\sigma^2 \end{pmatrix}. \end{aligned} \tag{5.28}$$

In summary, the analytical integral over $\mathbf{A}$ can be performed for each element in the SNR timeseries (corresponds to $\mathbf{J}$ vectors in the equations) and for each (α,δ), and produces a time series $I(t_c,\alpha,\delta)$. The posterior of (α,δ) needs one more numerical integral

over t_c:

$$p(\alpha, \delta \mid \mathbf{d}) \propto \int_{t_c-T}^{t_c+T} I(t_c, \alpha, \delta) dt_c. \tag{5.29}$$

5.2.4 Sky pixelation and confidence areas

In practice, we pixelate the sky and calculate the posterior probability density on each pixel. We adopt the adaptive HEALPix scheme [452, 459] to accelerate the calculation. We first divide the sky into $N_{\rm pix,0} = 3072$ pixels and calculate the posterior probability of each pixel through Eq. 5.29 and we assume the posterior probability is constant within a pixel, thus the probability for each pixel is equal to the calculated posterior probability multiplied by the area. The top $N_{\rm pix}/4$ most probable pixels are further divided into $N_{\rm pix}$ daughter pixels, and posterior probability is calculated again for those pixels. Repeating this, the finest resolution increases exponentially while the computational complexity is linear to the number of iterations.

After calculating posterior to the desired resolution, we define the *confidence areas* as follows. We rank the probability of all pixels in descending order and search from the first pixel. The probability of searched pixels is accumulating during the search, until the cumulative sum equals the given credible levels, e.g., 0.9 or 0.5. The area in that credible level is given by the sum of the area of searched pixels.

We also define a *searched area*. The searched area is computed by searching from the first pixel defined above until the true sky location is included. The searched area is the smallest of such constructed area that contains the true sky direction of the source. It measures the accuracy of the localization, while the confidence areas measure the precision.

5.2.5 Tests and implementation on current detectors

SealGW has been tested on the LHV network under O2 sensitivity and the design sensitivity. The area statistics, p-p plot, and comparisons with `LALInference` can be found in Ref. [4]. Note that in Ref. [4] SealGW uses a correlated prior for $\mathbf{A}$, while in this chapter we adopt the uncorrelated one.

SealGW has been written into a Python package[1] and uploaded to the Python Package Index (PyPI)[2]. It can be installed by

```
pip install sealgw
```

and an example-based tutorial is provided along with the package.

SealGW has been implemented into the detection pipeline SPIIR [300] and has been tested on O3 mock data along with SPIIR pipeline, and showed good consistency with

[1] https://github.com/marinerq/sealgw
[2] https://pypi.org/project/sealgw/

`Bayestar` in BNS and NSBH sources, although the skymaps for BBH were biased. A possible reason for the poor performance on BBH systems is that the prior of **A** requires a good fitting of a specific astrophysical model, which, in the sense of the range of chirp mass, is simple for BNS and NSBH sources, but becomes more difficult for BBH systems.

By implementing SealGW into a real pipeline, we do not aim to replace the current localization algorithm `Bayestar`, but want to provide an alternative method that requires fewer computation resources and brings a potential speed-up. For example, during O4, `Bayestar` has priority access to a large number of CPUs and it automatically downloads the detected events and analyzes them, during which Bayestar will use more than one hundred threads to achieve a subsecond performance. By contrast, SealGW can achieve subsecond performance with only 4 threads (details in Sec. 5.4). In a detection pipeline, it is possible to calculate and include the skymap at the initial submission of triggers since SealGW does not take up too much computational resources and can run on the detection node. Therefore, it saves the time cost of data transfer and `Bayestar` calculation, which brings a potential speed-up for the EM follow-up. However, this feature requires the Python3 version of the SPIIR pipeline, which is still under development or review at the time of writing this thesis.

5.3 Extending SealGW to long signals

5.3.1 Effects of the Earth rotation

The antenna response functions are functions of time due to the Earth's rotation. They are regarded as constants in CBC data analysis for current detectors because the signal durations are typically short, during which the Earth moves little. However, for long signals in the 3G detectors, the Earth rotation will induce modulations to the signal, as shown in Fig. 5.4. Converting GW frequency to time to the merger by Eq. 1.81, we find that the Earth rotation effects are visible when the signal duration is longer than 10 minutes.

The Earth's rotation needs to be taken into consideration in data analysis because it encodes more informative triangulation information by extending the baseline of the detector network [112, 460, 461]. Neglecting the Earth's rotation may have little impact on detection, as the SNR is mostly contributed by the last stages of the signal (see Fig. 5.5). However, ignoring it causes a loss of information and could lead to biases or inefficiency in parameter estimation, especially for sky location parameters.

The non-constant antenna response functions would result in a non-constant effective distance (Eq. 2.20), so the effective distance cannot be factored out, in other words, the matched filtering formalism we introduced in Sec. 2.1.1 does not hold anymore. In addition, direct matched filtering to long signals could be extremely computationally expensive. We

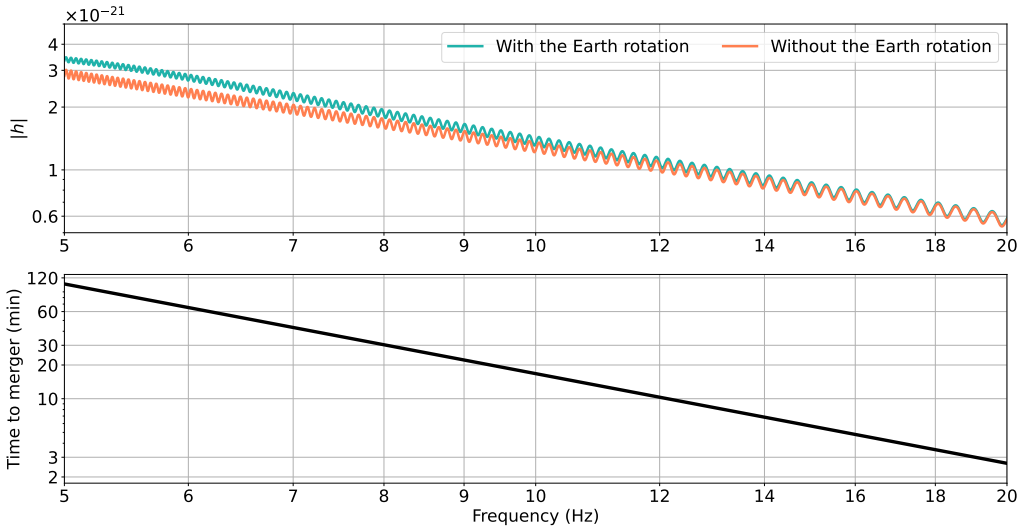


Figure 5.4: Illustration of the Earth rotation effects with a $1.44M_{\odot}$+$1.37M_{\odot}$ (detector frame mass) binary system. **Upper panel**: the frequency domain detector response calculated with (green) and without (orange) the Earth rotation. The small oscillations are due to spin precession. **Lower panel**: the time to the merger as a function of GW frequency.

need new methods for long signal detection.

5.3.2 Multi-banding

In this work, we demonstrate that multi-band analysis [289, 291, 443] is an effective way of solving the aforementioned issues, and fast localization algorithms can be built upon multi-band detection statistics.

Multi-band analysis is based on the fact that orbital evolution of the quasi-circular BNS inspiral stage is well modeled. Observable BNSs are not likely to have large spins [136, 462] or precession, and the frequency evolves as Eq. 1.82 to the Newtonian order. The monotonic evolution of GW frequency ensures a one-to-one correspondence between time segments and frequency bands, i.e., chopping the GW waveform into multiple time segments, $[t_n, t_{n-1}), [t_{n-1}, t_{n-2}), \ldots, [t_1, t_0)$, results in a corresponding sequence of frequency bands $[f_n, f_{n-1}), [f_{n-1}, f_{n-2}), \ldots, [f_1, f_0)$ defined via Eq. 1.82. One can choose the length of time intervals such that within each interval the Earth's rotation can be ignored, i.e. the detectors' antenna response functions can be assumed constant and the current matched filtering techniques can be directly employed. One can also down-sample the data in each frequency band according to its highest frequency to reduce computational cost. Results from each time segment can be combined in succession as new data comes in.

Fig. 5.5 shows the multi-band scheme in this work. We consider the negative latency

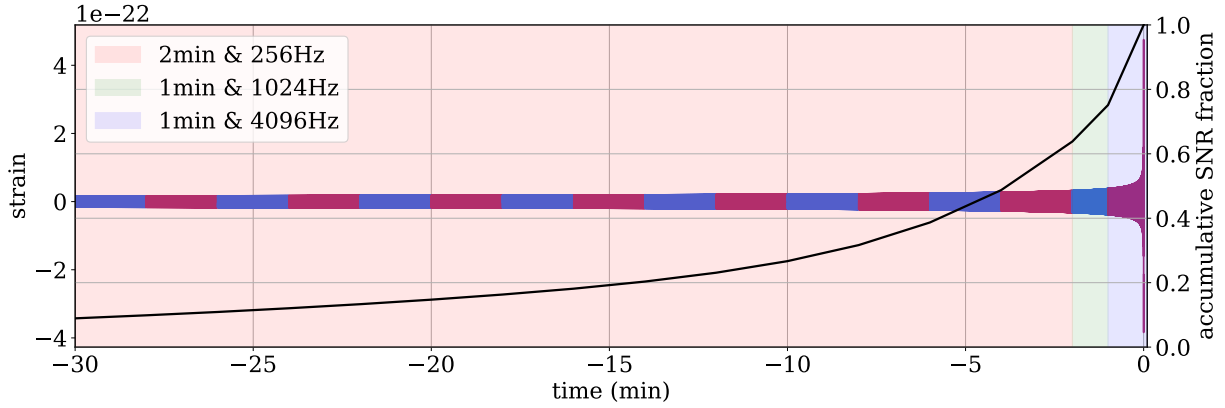


Figure 5.5: Multi-banding scheme for this work. **Left axis:** An illustration of a chopped GW waveform that is alternately colored for different waveform bins and sampling frequencies. We use two-minute segments with 256 Hz sampling frequency for waveforms from 60 minutes before the merger to two minutes before the merger, and one-minute segments for the last two minutes with 1024 Hz and 4096 Hz sampling frequencies, respectively. **Right axis:** The cumulative SNR of the signal, showing the contribution of SNR from each time segment.

up to 60 min, and choose 2 min segments and 256 Hz sampling frequency until the final two minutes. In the last two minutes, the SNR grows rapidly, and the GW reaches high frequencies, therefore a finer time resolution is used to improve the detection and localization. In addition to the limit from Nyquist-Shannon sampling theorem, in practice, we find that a sampling rate that is higher than Nyquist frequency could be helpful in localizing high SNR events. As a demonstration, we equally divide the last two minutes, and employ sampling frequencies of 1024 Hz and 4096 Hz, respectively. A more elaborate segmentation is also sensible e.g. [463], but we leave a comprehensive investigation of the multi-band scheme to future works.

Filtering each timeseries in the multi-band scheme produces a separate complex SNR timeseries that must be coherently combined to achieve a precise localization [452, 464]. The gain in precision for long signals comes not only from the accumulation of absolute SNR, but also from phase drifts and amplitude modulations of the SNR timeseries due to the Earth's rotation. A direct combination (linear addition) of multiple SNR timeseries is feasible for short signals, as it only requires a set of combination parameters to align the template bands in time and phase [291]. However, for long signals, the direct combination scheme is no longer coherent because of the phase drifts due to the Earth's rotation. Including the phase drifts in combination parameters can lead to a coherent addition, but it loses information contained in the changing time delays between detectors which depends on the sky location. Therefore, instead of directly adding SNR, we multiply likelihoods from every band before marginalization over nuisance extrinsic parameters. Since there is little overlap between bands (contributed by noise correlation and template

overlaps which are windowed out), they can be treated as independent measurements, as if there are different detectors at different frequency bands.

For each band, we simulate signals and fit the prior coefficients for SealGW. We consider a three-detector network with one triangle ET at the Virgo site and two L-shaped CEs at the H1 and L1 sites, respectively. The fitting coefficients are given in Table 5.1. If multiple bands are combined, the coefficients of the latest band will be employed as it contains the highest SNR and should be more accurate.

Having the prior assigned and likelihood combined, we can perform multi-banding analysis for the long signals and update the skymap as new data come in. Thanks to the efficiency of SealGW, adding more bands does not bring too much computational burden, and the results are given in the next section.

5.4 Tests on mock catalog

5.4.1 Catalog simulation

To assess the performance of the above localization scheme, we simulate a mock BNS catalog with the same population model used in Chapter 3 for BNS (Eq. 3.11). We set $\mathcal{R}_{\rm obs}(z=0) = 320{\rm Gpc}^{-3}{\rm yr}^{-1}$, i.e., the medium merger rate catalog used in Chapter 3. We simulate 68000 BNS sources within $z = 3$, corresponding to roughly one month of observations. We assume neutron star mass is uniformly distributed in $[1.1M_\odot, 2M_\odot]$ in the source frame and isotropic sky distribution and inclination. Note that the location and configuration of detector networks are not settled yet and subject to change, and the BNS mass distribution and the current merger rate density estimate have large uncertainties due to the as yet small number of BNS detections.

Signals are injected into Gaussian noise realizations and analyzed individually, i.e., we do not consider them to be overlapping with each other. Overlapping signals could cause dominant biases in parameter estimation, but mainly in the case when the merger times are very close [1, 375, 376, 377, 465, 466]. Among the 68000 BNS events evenly distributed in one month, roughly 1.3% of them have another event ending $< 0.5\,$s afterwards. Many of the signals are not actually detectable, further reducing the chance of significant bias. Even though the number of overlapping signals could be large given the large number of events expected to be detected with 3G detectors, our simulation will still apply to the vast majority of non-interfering signals.

We use the waveform model `TaylorF2` [207, 467, 468] to generate GW signals and map frequency to time before merger via the stationary phase approximation with Eq. 1.81. The 3.5 post-Newtonian waveform is a reasonable choice for analyzing quasi-circular inspiralling compact binaries [469]. Several works suggest some non-quasi-circular binaries, like precessing and eccentric systems, or systems with strong higher order emission, can

	a	b	c	d
60 to 58 min	1.461E-04	4.154E-04	9.775E-05	1.983E-04
58 to 56 min	1.422E-04	3.058E-04	6.771E-05	6.294E-04
56 to 54 min	1.255E-04	4.906E-04	1.059E-04	-1.055E-04
54 to 52 min	1.245E-04	4.195E-04	8.435E-05	2.100E-04
52 to 50 min	1.229E-04	3.204E-04	7.251E-05	3.448E-04
50 to 48 min	1.144E-04	3.293E-04	5.848E-05	4.842E-04
48 to 46 min	1.104E-04	2.823E-04	7.745E-05	8.656E-05
46 to 44 min	1.090E-04	2.290E-04	7.410E-05	9.308E-05
44 to 42 min	1.035E-04	2.354E-04	5.456E-05	3.028E-04
42 to 40 min	1.168E-04	-2.682E-04	5.740E-05	3.593E-04
40 to 38 min	8.510E-05	3.364E-04	5.744E-05	1.452E-04
38 to 36 min	8.702E-05	2.223E-04	5.182E-05	2.044E-04
36 to 34 min	8.659E-05	1.585E-04	4.910E-05	1.666E-04
34 to 32 min	8.490E-05	9.356E-05	4.102E-05	2.373E-04
32 to 30 min	8.114E-05	6.739E-05	3.843E-05	2.354E-04
30 to 28 min	7.627E-05	6.933E-05	4.238E-05	1.044E-04
28 to 26 min	6.951E-05	1.009E-04	4.521E-05	2.268E-05
26 to 24 min	6.952E-05	1.159E-05	3.532E-05	1.395E-04
24 to 22 min	6.336E-05	3.479E-05	2.678E-05	2.099E-04
22 to 20 min	5.454E-05	8.247E-05	3.584E-05	6.971E-06
20 to 18 min	5.162E-05	5.141E-05	2.508E-05	1.447E-04
18 to 16 min	4.474E-05	7.013E-05	2.915E-05	1.150E-05
16 to 14 min	3.921E-05	5.999E-05	2.052E-05	1.032E-04
14 to 12 min	3.377E-05	5.805E-05	1.929E-05	6.461E-05
12 to 10 min	2.776E-05	4.991E-05	1.445E-05	8.852E-05
10 to 8 min	2.193E-05	4.649E-05	1.041E-05	9.040E-05
8 to 6 min	1.584E-05	5.093E-05	9.336E-06	5.358E-05
6 to 4 min	4.452E-06	1.543E-04	1.335E-05	-6.609E-05
4 to 2 min	7.450E-06	3.545E-05	4.067E-06	4.058E-05
2 to 1 min	7.450E-06	3.545E-05	4.067E-06	4.058E-05
1 to 0 min	3.820E-06	1.897E-05	2.376E-06	1.728E-05

Table 5.1: Prior coefficients for the SealGW multi-banding analysis. The prior coefficient μ and σ follows $\mu = a\text{SNR} + b$ and $\sigma = c\text{SNR} + d$, and a, b, c, d are given in the table. Since the relation is fit with detectable events, the sources that are detectable at very early time are expected to have small luminosity distance (and therefore larger A_{ij}). This expectation is consistent with the tendency in the table that earlier bands have larger a (a determines the mean value of A_{ij}).

be better localized [470, 471, 472, 473, 474]. However, that would require novel search algorithms (e.g. [288]) upon which new fast localization methods would have to be built, because current localization methods, including `Bayestar` and SealGW , are based on aligned-spinning waveform templates in which plus and cross polarizations of GWs only have a phase difference.

We perform matched filtering assuming a perfect knowledge of intrinsic parameters and set total SNR>12 as the detection criterion, where total SNR is converted from the multi-band matched filtering outputs by the analytical expression of SNR at Newtonian order [289]. Matched filtering with known injection parameters is the ideal case, while in a realistic scenario, one should build a template bank that achieves a reasonable match (e.g. >97%) everywhere in the parameter space. The purpose of this work is to assess the performance of the multi-band localization scheme. We leave a dedicated long signal early warning pipeline and simulations with more realistic mock data for future work.

5.4.2 Localization statistics

For each simulation, we perform multi-band matched filtering from 60 minutes before the merger with low frequency cutoff at 5 Hz. Fig. 5.6 shows the cumulative number of events for different negative latencies. ~ 10 events can be localized within 100 deg^2 20 minutes before the merger, and 6 minutes before the merger the number of events increases to ~ 100. Also, $\sim 1-10$ events can be localized within 10 deg^2 up to 6 minutes before merger.

Extreme early warnings are possible. Several events in our simulation are detected and preliminarily localized 40-50 minutes before the merger and this number could be underestimated since our analysis has hard cutoffs at 5 Hz and one-hour negative latency. ET would be able to collect sensible data down to ~ 3 Hz and trigger even earlier detections [167, 171]. However, BNS with high negative latencies is not likely to be well localized until more data comes in, bringing higher SNRs, wider frequency bands and a longer equivalent network baseline. Multi-band analysis helps update detection statistics and skymaps on the fly. Fig. 5.7 shows the evolution of skymaps and localization areas in our simulation. The area evolution agrees with the theoretical prediction that the localization area is inversely proportional to the SNR and effective frequency bandwidth [454]. The example skymap is from a $1.4+1.4M_\odot$ BNS at 1000 Mpc detected 30 minutes before the merger. It presents nested contours with new bands combined in succession and is finally pinpointed within 0.2 deg^2, but is already well localized ~10 minutes before the merger. The localization area traces in the lower panel show the decreasing rate of localization areas: those localized within 100 deg^2 ~ 20 minutes before the merger in Fig. 5.6 are generally detected 40-50 minutes before the merger.

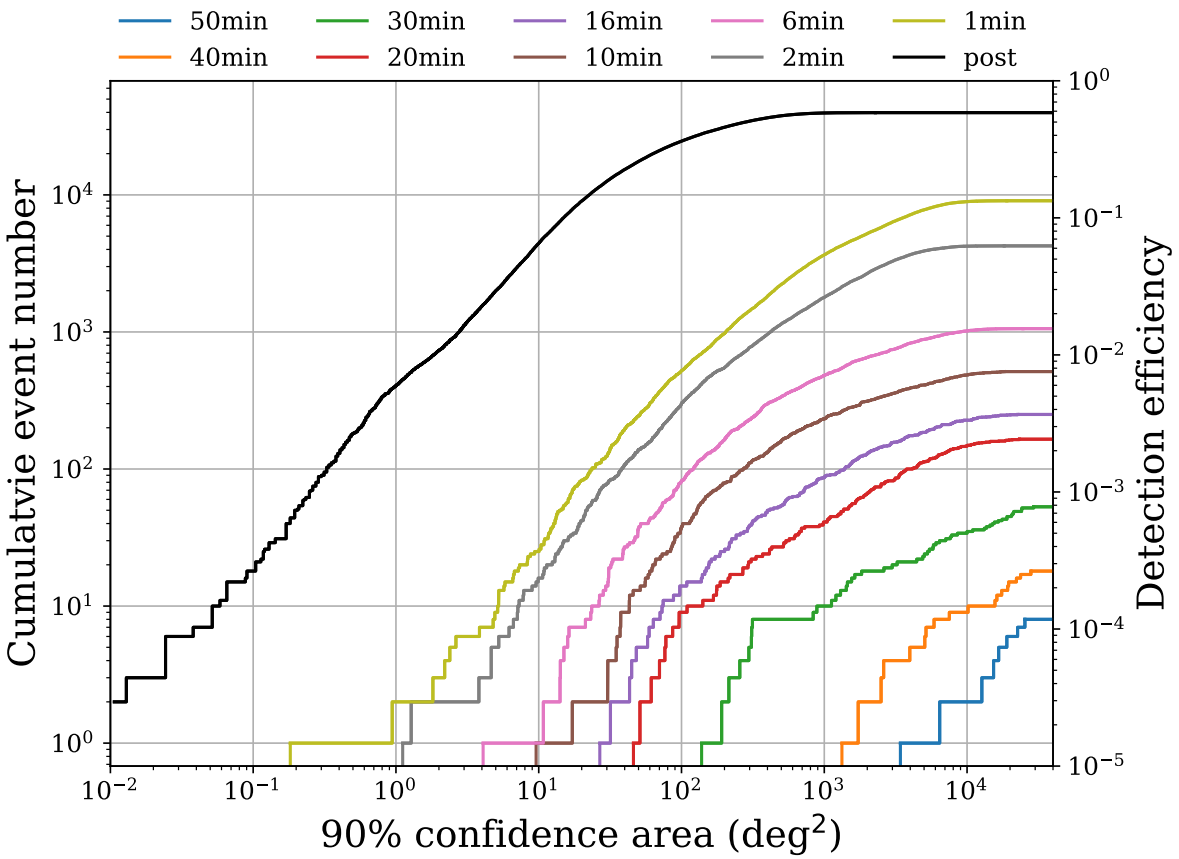


Figure 5.6: Cumulative number of detections and 90% confidence sky localization areas for the 68000 BNS simulations (roughly one month of observation). The corresponding detection efficiency is labeled in the right y axis. Note that here detection efficiency of 100% means detecting all sources within $z < 3$. We choose 10 different negative latencies (from 50 minutes to post-merger) and the curves show the cumulative distribution of 90% areas of events that are detected at those times.

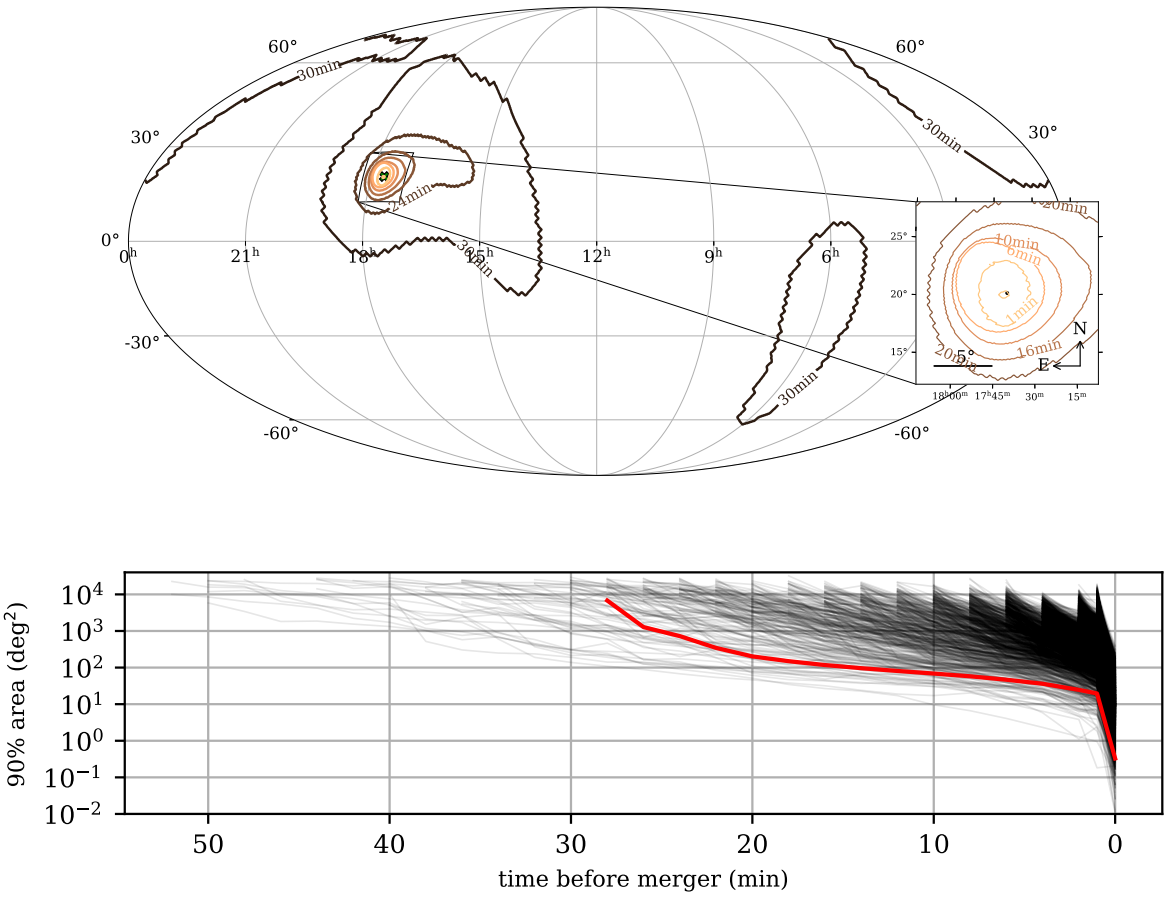


Figure 5.7: Skymap evolution. **Upper panel**: An example skymap for a $1.4 + 1.4 M_{\odot}$ BNS at 1000 Mpc detected 30 minutes before merger with a network SNR of 12. The SNR increases to 17 at 20 minutes before the merger, 31 at 10 minutes, 95 at one minute and 130 after the merger. We show the 90% localization contours at different negative latencies. The injection sky location is marked with a cross. **Lower panel**: Evolution of 90% confidence localization areas of early warning events in our simulation. The example in the upper panel is plotted in red line. The localization is performed when the source is detected and the detection is updated every two minutes except for the last two minutes. Therefore, sets of lines show up every two minutes.

5.4.3 Diagnoses

P-P plot

Fig. 5.8 is the P-P plot of our localization simulation, showing x% confidence region (x-axis) is able to include y% of total events (y-axis, scaled). The diagonal shapes show the multi-band localization scheme is reasonably self-consistent. The lines for 30+ minutes before the merger have larger statistical fluctuations due to the insufficient number of samples.

Time cost

We tested the time cost of SealGW and `Bayestar` calculation with the same data (full-bandwidth ET+2CE network) and skymap resolution ($n_{side} = 2048$, finest pixel = 0.0008 deg^2), as shown in Fig. 5.9. Tests are performed on a 2.44 GHz processor with OpenMP multithreading. Thanks to the semi-analytical property, SealGW can achieve ~ 26 times faster speed than `Bayestar` with fewer threads, and the speed-up factor goes down to ~ 4 when more threads come in as the non-parallelizable calculation begins to dominate SealGW run time. It only takes SealGW $\sim$ 3s with 1 thread and $\sim$ 0.5s with 8 threads, which means SealGW is able to perform real-time localization with a low hardware requirement. Note that the time cost can be further reduced with narrower bandwidth or coarser skymap resultions, e.g., the time cost of SealGW can be halved when the finest resolution is 0.013deg^2. The efficiency and cheapness are suitable for the 3G detector scenario in which the number of detections can be huge. Nevertheless, a thorough estimate of early warning latency would require a comprehensive design of detection pipeline structure, and there would be a wall time of $\sim$ 0.1s to read and preprocess the data from pipeline outputs.

5.5 Summary and discussions

We provide an exploratory demonstration of early warning localization of long signals for 3G GW detector networks. We simulate a mock catalog for one month of observation with an ET+2CE network, and perform multi-band analysis with the fast localization algorithm SealGW . We show that this is an efficient scheme for pre-merger localization.

Multi-band analysis allows us to detect BNS in an early stage and update the results regularly with incoming data. There are tens of BNS detected more than 30 minutes before the merger in our simulations, and localized within 100 deg^2 at $\sim$ 10 minutes before the merger. 10 deg^2 can be achieved $\sim$ 6 minutes before merger. Since wide-field optical transient facilities usually have a field of view of 1-10 deg^2 (see summaries in [437, 449]), the precise pre-merger localization of BNS would be extremely helpful to finding EM

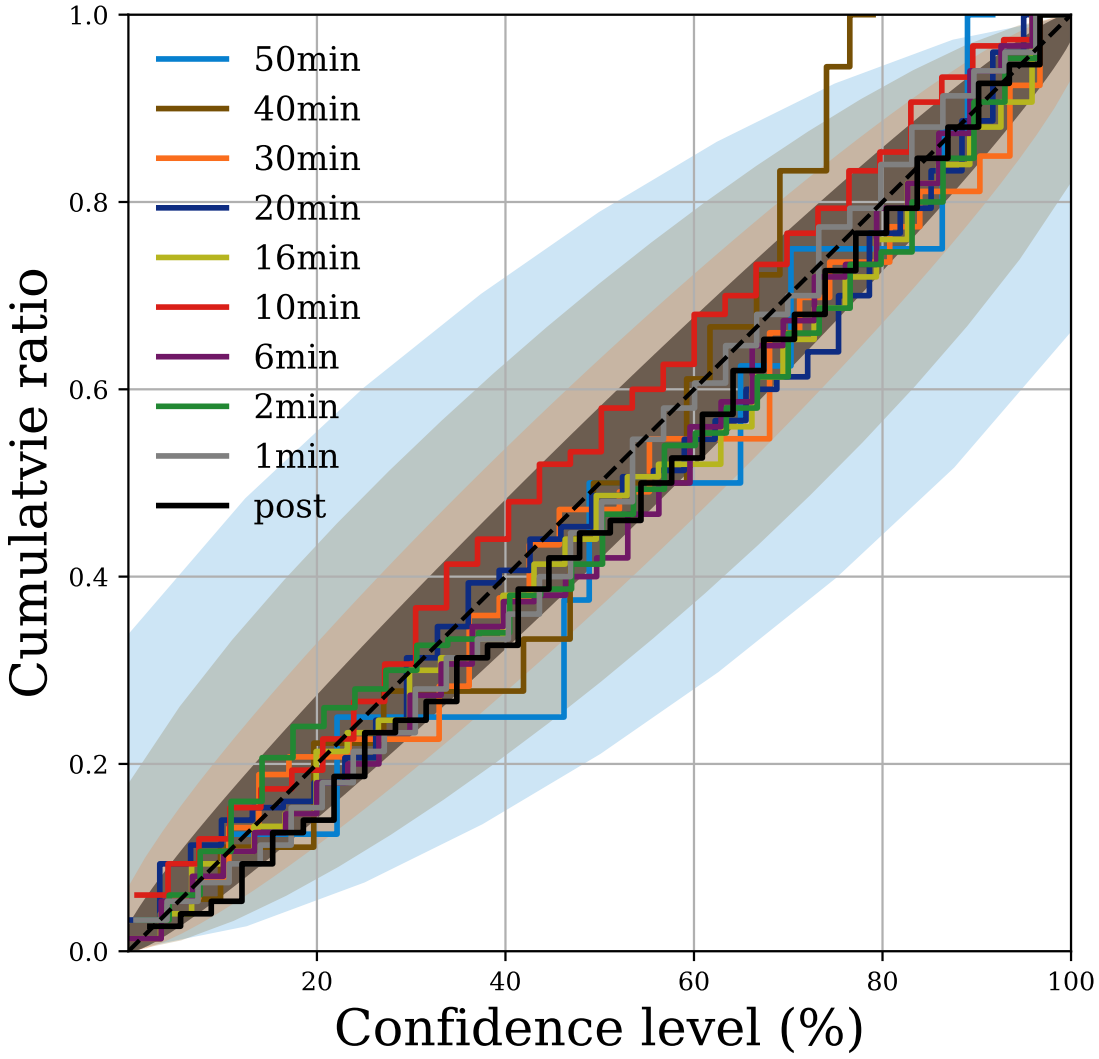


Figure 5.8: P-P plot of localizations in our simulation at different bands. For 50 min (light blue, sample size = 8), 40 min (brown, sample size = 18), and 30 min (orange, sample size = 53), error bars are plotted individually with their own colors. For other bands, we randomly select 150 events and plot their error bar in black. The error bar is calculated from a binomial distribution, and we note that it only converges to (0%, 0) and (100%, 1) when the sample size is sufficiently large.

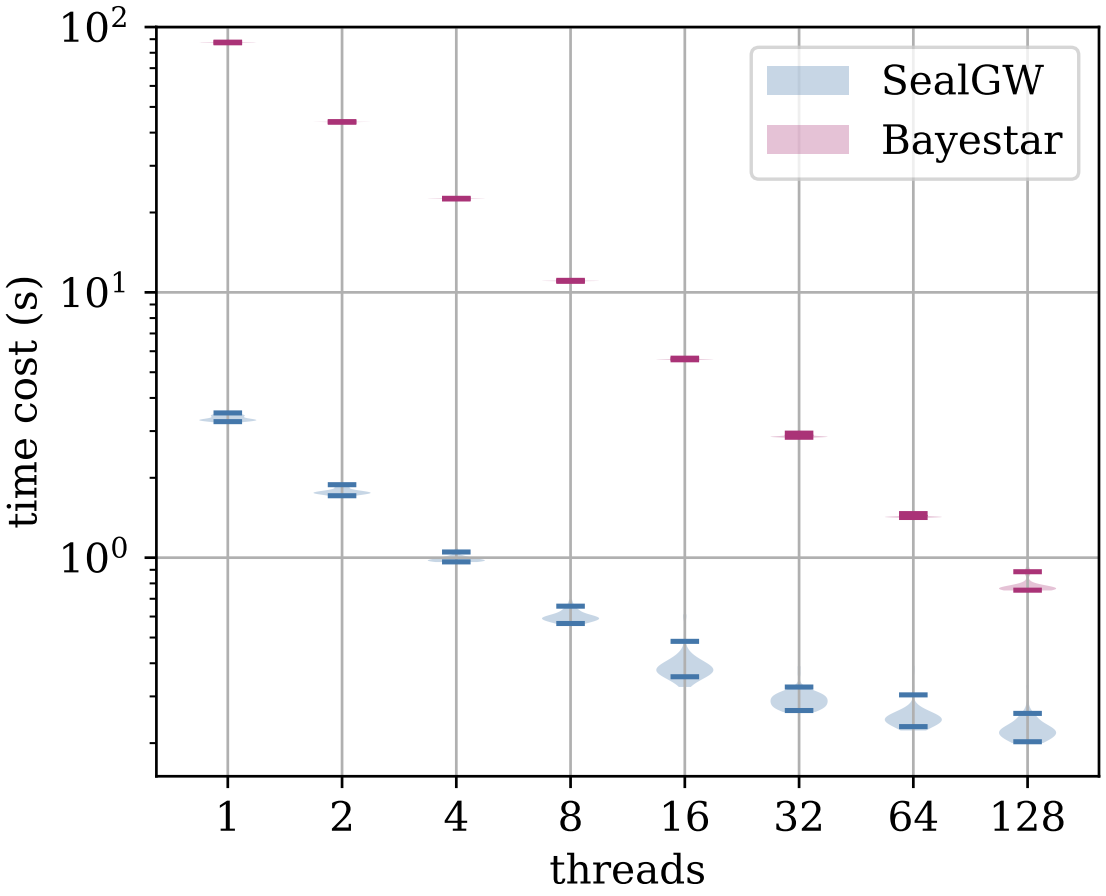


Figure 5.9: Time cost of running SealGW and `Bayestar` for ET+2CE network on a 2.44 GHz processor with different number of threads, excluding the time costs of matched filtering and data conditioning. The two algorithms are tested with the same data, and skymaps are calculated to the same level of resolution ($n_{side} = 2048$, finest pixel = 0.0008 deg^2). The matched filtering speed (detection latency) depends on the efficiency of the detection pipeline, and data conditioning usually takes ~ 0.1s to read and pre-process the data from pipeline outputs.

counterparts before the merger and observing the entire process of BNS coalescence.

Our work here presents a solution for the crucial step of performing real-time localization in the context of online searches in 3G detectors, that effectively reduces the latency and computational burden arising from pre-merger localization. However, there remains the larger issue of developing the surrounding infrastructure to search for pre-merger signals and disseminating sky-maps in low latency to observatories before detection. As an exploratory demonstration, we have made several simplifications to the problem, such as ignoring overlapping signals, assuming perfect matched filtering, and a relatively naive waveform segmentation. The merger rate estimation of BNS is also uncertain to date, therefore the absolute detection numbers should be interpreted as an order-of-magnitude estimation. We plan to explore the multi-band analysis on a real detection pipeline and use a more accurate astrophysical population (which should be available in years with new observations) in our future work.

The skymap fits files for 30, 20, 10, 6 and 1 minutes early warning and post-merger triggers are openly available in zenodo [475].

Chapter 6

Rapid full parameter estimation of BNS in the 3G era

Having demonstrated the fast sky localization of long BNS signals, we now advance to a more challenging problem: the full parameter estimation of long BNS signals for the 3G detectors. Traditional stochastic sampling methods have struggled with this task due to the signal length, Earth's rotation effects, and high SNR, making them prohibitively slow. In this chapter, we present a solution using neural density estimation combined with a series of data pre-processing techniques. More specifically, we develop an efficient multi-banding scheme that significantly compress the length of data, use heterodyning to reduce the complexity of data, and employ singular value decomposition and a deep neural network to further compress the data, which is used to condition a normalizing flow to produce density estimation. By comparing our models with Fisher matrix forecasts and SealGW, we demonstrate that our approach can precisely estimate the BNS parameters while maintaining decent accuracy and self-consistency within a reasonable parameter space. Our neural density estimator can draw 5000 posterior samples within 0.3s, showcasing its potential for application to a large number of events in next-generation detectors.

This chapter is organized as follows. Sec. 6.1 introduces the generic idea and network structure of neural density estimation and Sec. 6.2 introduces normalizing flows, the generative model we use for neural density estimation. We describe how the data is compressed in Sec. 6.3 and how the model is built and trained in Sec. 6.4. We show the results in Sec. 6.5 and summarize and discuss the results in Sec. 6.6.

6.1 Neural networks for parameter estimation

Full parameter estimation with stochastic sampling is computationally intensive, with time costs increasing significantly as the SNR and signal length grow. In the 3G era, the GW event rate will rise sharply due to improved detector sensitivity, and in Sec. 2.4 we have conservatively estimated that it could take tens of millions of CPU hours to analyze events in one-year observation if we use the traditional stochastic sampling methods, placing a substantial burden on researchers, funding agencies, and even contributing to global warming. On top of that, some analyses for high SNR or long signals, or with slow waveforms, are beyond the reach of traditional methods.

Neural density estimation aims to train a neural network to produce a probability density conditioned on the data $q(\theta|d)$ to approximate the true posterior distribution $p(\theta|d)$. Once trained, these models could produce accurate posterior samples within seconds, as demonstrated in `Vitamin` [356] and `DINGO` [357]. The samples can be further refined by importance sampling [358] in minutes. The speed of neural density estimation offers significant advantages in processing large datasets and handling complex analyses, especially in the context of the 3G era of GW detection. For example, a recent work [142] systematically explored eccentricity in CBC systems on the population level using `DINGO`, which would be prohibitively slow using traditional methods due to the computational demands of eccentric waveform models.

Neural density estimation relies on generative models, a class of machine learning models that could learn the underlying distribution of data. Various generative models have been applied to GW parameter estimation, including variational autoencoders [356] and normalizing flows [357]. Normalizing flows have emerged as the preferred choice because it directly learns and optimizes the likelihood function, which is essential for accurate parameter estimation. In contrast, likelihood is often intractable and must be approximated in other generative models.

Although neural network structures are flexible, flow-based density estimators have two key modules that all models should have: an embedding layer and a conditioned normalizing flow. The former compresses the data to a more compact form to reduce the size of the model and the latter takes the compressed data for inference. The general structure is shown in Fig. 6.1 and detailed explanations will be given in the next two sections.

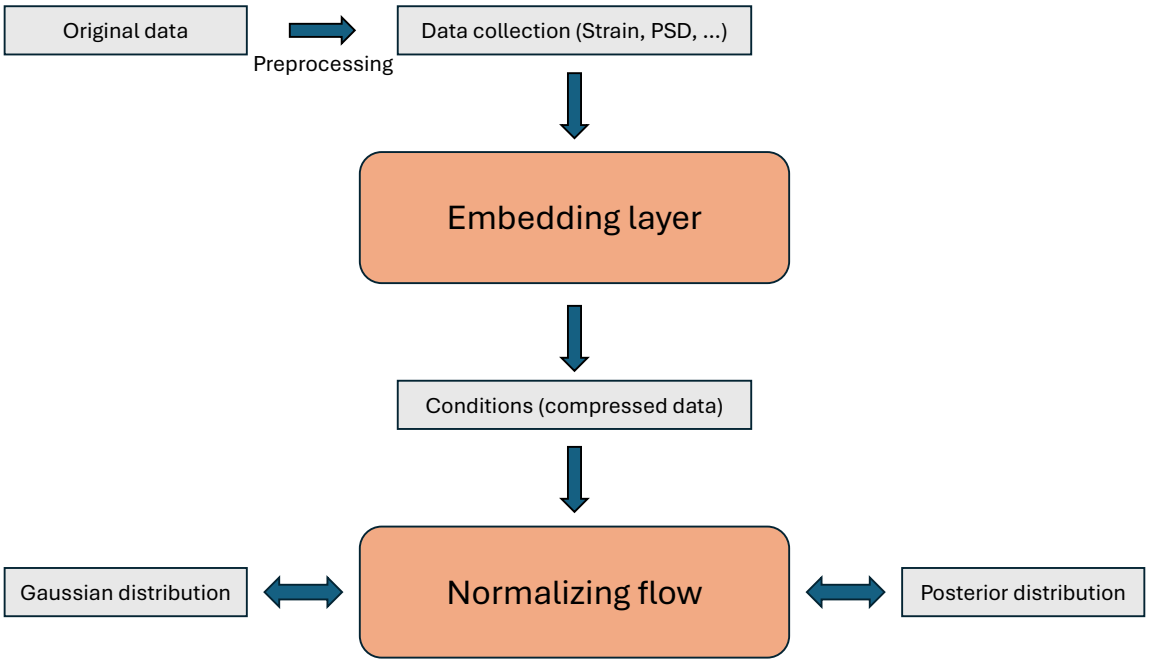


Figure 6.1: General structure of flow-based neural posterior estimators. It consists of an embedding layer that compresses the data on which a normalizing flow is conditioned. The normalizing flow will learn transforms between the Gaussian distribution and the target posterior distribution.

6.2 Normalizing flows

6.2.1 Basic principles

Normalizing flow is a type of generative machine learning model that learns the distribution $p_{\mathcal{X}}(\boldsymbol{x})$ of data space $\mathcal{X}$ [476, 477]. A normalizing flow is composed of a series of learnable and invertible transformations f_i that maps the data space $\mathcal{X}$ to a latent space $\mathcal{Z}$ (i.e., transforming $\boldsymbol{x}$ to $\boldsymbol{z}$), in which variable $\boldsymbol{z}$ follows a simple distribution such as Gaussian. The transformations need to be *invertible* so that it allows us to sample $\boldsymbol{z}$ in the latent space easily and transform it to $\boldsymbol{x}$ to approximate the complex probability density in the data space. An illustration of this process is shown in Fig. 6.2.

The mathematical foundation of normalizing flows is the rule of change of variables in probability theory. Let $f = f_k \circ f_{k-1} \circ \dots f_1$ where f_i are the transformations shown in Fig. 6.2, and assume f is parametrized by the trainable parameters $\boldsymbol{\phi}$, we have $\boldsymbol{x} = f(\boldsymbol{z_0}; \boldsymbol{\phi})$. We further require the transforms to be *bijective* and *differentiable*. Let $g = f^{-1}$ and $g_i = f_i^{-1}$, we have

$$p_{\mathcal{X}}(\boldsymbol{x}) = p_{\mathcal{Z}}(g(\boldsymbol{x}; \boldsymbol{\phi})) \left| \det \left(\frac{\partial g(\boldsymbol{x}; \boldsymbol{\phi})}{\partial \boldsymbol{x}} \right) \right|, \tag{6.1}$$

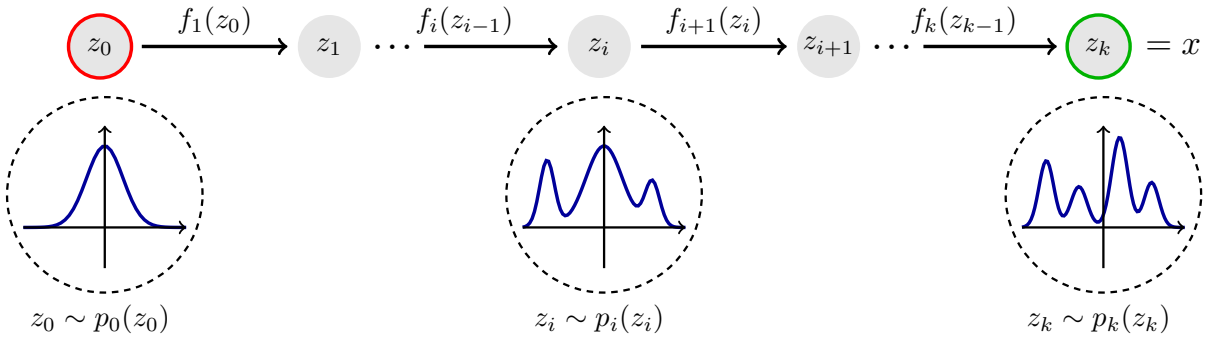


Figure 6.2: Illustration of normalizing flows. z_0 follows the Gaussian distribution, and after a series of transformations f_i, it becomes z_k whose distribution approximates the target distribution of data x. Image credit: Awesome Normalizing Flows.

where

$$\left|\det\left(\frac{\partial g(\boldsymbol{x};\boldsymbol{\phi})}{\partial \boldsymbol{x}}\right)\right| = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_n} \end{vmatrix} \tag{6.2}$$

is the absolute value of the Jacobian determinant of the transformation g, which we always have access to because g is differentiable. We can also calculate $p_{\mathcal{Z}}(g(\boldsymbol{x};\boldsymbol{\phi}))$ because z is constructed to be a trivial distribution. Therefore, we have an analytical expression of the data distribution $p_{\mathcal{X}}(\boldsymbol{x})$ in normalizing flows. This contrasts with other generative algorithms, in which the probability distribution is not explicitly known.

6.2.2 Constructing transformations

To construct a normalizing flow, the mapping f must be invertible and its Jacobian must be tractable. There are a number of methods for constructing normalizing flows that satisfy these two requirements. As an example, I introduce *coupling flows* [478] in this chapter, including some common variants of coupling flows.

Coupling flows split D-dimensional input data $\boldsymbol{x} \in \mathbb{R}^D$ into two subspaces $\boldsymbol{x} = (\boldsymbol{x_a}, \boldsymbol{x_b}) \in \mathbb{R}^d \times \mathbb{R}^{D-d}$. For each transformation, $\boldsymbol{x_a}$ is transformed by an identity mapping, while $\boldsymbol{x_b}$ is mapped by a function $\mathcal{G}(\cdot;\boldsymbol{\vartheta})$ parametrized by $\boldsymbol{\vartheta}$, which further depends on $\boldsymbol{x_a}$

$$\boldsymbol{\vartheta} = \Theta(\boldsymbol{x_a}). \tag{6.3}$$

In summary, the transformation can be written as

$$f(\boldsymbol{x}) = \boldsymbol{y} = \begin{cases} \boldsymbol{y}_a = \boldsymbol{x}_a \\ \boldsymbol{y}_b = \mathcal{G}\left(\boldsymbol{x}_b; \Theta\left(\boldsymbol{x}_a\right)\right) \end{cases} \tag{6.4}$$

The two subspaces are coupled via the *conditioner* $\Theta(\boldsymbol{x_a})$. Θ can be parametrized by pa-

rameters $\boldsymbol{\phi}$, which need to be optimized during the training. The coupling transformation is easily invertible as long as $\mathcal{G}$ is invertible:

$$f^{-1}(\boldsymbol{y}) = \boldsymbol{x} = \begin{cases} \boldsymbol{x}_a & = \boldsymbol{y_a} \\ \boldsymbol{x}_b & = \mathcal{G}^{-1}\left(\boldsymbol{y}_b; \Theta\left(\boldsymbol{x_a}\right)\right). \end{cases} \tag{6.5}$$

Most coupling functions are chosen to be element-wise: the i'th element in $\boldsymbol{y_b}$ only depends on the i'th element in $\boldsymbol{x_b}$ and the parameters $\boldsymbol{\vartheta_i} = \Theta_i(\boldsymbol{x_a})$, i.e.,

$$\begin{aligned} \boldsymbol{y}_b &= (\mathcal{G}_1(\boldsymbol{x_b}_1; \boldsymbol{\vartheta_1}), \ldots, \mathcal{G}_{D-d}(\boldsymbol{x_b}_{D-d}; \boldsymbol{\vartheta_{D-d}})) \\ &= (\mathcal{G}_1(\boldsymbol{x_b}_1; \Theta_1(\boldsymbol{x_a})), \ldots, \mathcal{G}_{D-d}(\boldsymbol{x_b}_{D-d}; \Theta_{D-d}(\boldsymbol{x_a}))) \,. \end{aligned} \tag{6.6}$$

By enforcing this, the Jacobian becomes a lower triangular matrix:

$$J = \begin{bmatrix} \mathbf{1}_{d\times d} & \mathbf{0}_{d\times(D-d)} \\ \frac{\partial \boldsymbol{y}_b}{\partial \boldsymbol{x}_a} & \frac{\partial \boldsymbol{y_b}}{\partial \boldsymbol{x_b}}, \end{bmatrix} \tag{6.7}$$

where $\frac{\partial \boldsymbol{y_b}}{\partial \boldsymbol{x_b}}$ is a diagonal matrix of shape $(D-d)\times(D-d)$. The determinant is

$$\det J = \det \frac{\partial \boldsymbol{y_b}}{\partial \boldsymbol{x_b}} = \prod_{i=d}^{D} \frac{\partial \boldsymbol{y}_i}{\partial \boldsymbol{x}_i} = \prod_{i=1}^{D-d} \frac{\partial \mathcal{G}_i(\boldsymbol{x_b}_i; \Theta_i(\boldsymbol{x_a}))}{\partial \boldsymbol{x_b}_i}. \tag{6.8}$$

The time complexity of calculating the determinant is $\mathcal{O}(N^3)$, but by constructing the lower diagonal Jacobian, the complexity is reduced to $\mathcal{O}(N)$. One coupling transform only changes one subspace of the data. Therefore, it is common to stack multiple coupling transforms which divide the data $\boldsymbol{x}$ differently to build a normalizing flow. For example, one can equally divide $\boldsymbol{x}$ into two subspaces and transform them alternately.

$\mathcal{G}_i$ and the conditioner are the key parts of a coupling flow. The conditioners can be any complex but well-behaved functions because their derivatives and inverses are not required. In practice, they are often represented by neural networks such as MLP and ResNet. The choice of $\mathcal{G}_i$ needs to satisfy the requirements of the flow. A simple choice is the *affine transformation*:

$$\mathcal{G}(x_i) = \lambda_i x_i + \mu_i, \;\; \lambda_i > 0, \tag{6.9}$$

where λ_i, μ_i are outputs of the conditioner. In this case, the Jacobian determinant reduces to the product of all λ_i which avoids calculating the derivatives. This type of transformation is known as Real Non-Volume Preserving (RealNVP).

To improve the expressivity of the flow, more complex $\mathcal{G}_i$ are proposed, including polynomial splines [479], cubic splines [480], and neural splines [481] and so forth. Among these, the Neural Spline Flow (NSF) [481] has achieved the state-of-the-art thanks to its

flexibility and expressivity. The NSF is built upon monotonic rational-quadratic splines, which take the form of a quotient of two quadratic polynomials. The splines can be defined by the values and the derivatives of the spline at the knots. The spline values and the derivatives are outputs of neural networks (conditioners) and the number and the range of the knots are customizable. The overall function is very complex since it contains a number of neural networks, but is easily differentiable and analytically invertible in the meantime. In this chapter, we will use NSF for our purpose. NSF is implemented in the `Python` package `glasflow` [482], which is based on `nflows` [483].

6.2.3 Conditional normalizing flows

For GW parameter inference, we want to approximate a conditional probability distribution $p(\boldsymbol{\theta}|\boldsymbol{d})$ because the posterior is conditioned on the data. The normalizing flow framework can be extended to *conditional normalizing flows* for this purpose. Let $\boldsymbol{c} \in \mathcal{C}$ be the conditions, conditional normalizing flows learn

$$p_{\mathcal{X}|\mathcal{C}}(\boldsymbol{x} \mid \boldsymbol{c}) = p_{\mathcal{Z}}(g(\boldsymbol{x}; \boldsymbol{\phi}, \boldsymbol{c})) \left| \det \left(\frac{\partial g(\boldsymbol{x}; \boldsymbol{\phi}, \boldsymbol{c})}{\partial \boldsymbol{x}} \right) \right| . \tag{6.10}$$

Implementing the conditions is trivial. For coupling flows, the conditions are implemented by including them in the conditioner, i.e., the conditioner will be a function of both $\boldsymbol{x_a}$ and $\boldsymbol{c}$:

$$f(\boldsymbol{x}) = \boldsymbol{y} = \begin{cases} \boldsymbol{y}_a = \boldsymbol{x}_a \\ \boldsymbol{y}_b = \mathcal{G}\left(\boldsymbol{x}_b; \Theta\left(\boldsymbol{x}_a, \boldsymbol{c}\right)\right) \end{cases} \tag{6.11}$$

For instance, if the unconditioned flow has an MLP conditioner that takes as input of $\boldsymbol{x_a}$, the conditioner of the corresponding conditioned flow simply takes as input of the concatenate vector $(\boldsymbol{x_a}, \boldsymbol{c})$. Since $\boldsymbol{c}$ is not transformed, all calculations of conditioned flows can be performed just as unconditioned flows, except that the conditioners take an extra input $\boldsymbol{c}$.

6.2.4 Loss functions

Let $\boldsymbol{\phi}$ be the trainable parameters of a normalizing flow, training the flow involves minimizing a loss function $L(\boldsymbol{\phi})$. It is common to use the forward KLD between the target distribution $p^*(\boldsymbol{x})$ and the the approximate distribution $p_{\mathcal{X}}(\boldsymbol{x} \mid \boldsymbol{\phi})$ as the loss function:

$$\begin{aligned} L(\boldsymbol{\phi}) &= D_{\mathrm{KL}}\left(p^*(\boldsymbol{x}) \| p_{\mathcal{X}}(\boldsymbol{x} \mid \boldsymbol{\phi})\right), \\ &= \mathbb{E}_{p^*(\boldsymbol{x})}\left[\ln p^*(\boldsymbol{x})\right] - \mathbb{E}_{p^*(x)}\left[\ln p_{\mathcal{X}}(x \mid \boldsymbol{\phi})\right], \\ &= \mathbb{E}_{p^*(\boldsymbol{x})}\left[\ln p^*(\boldsymbol{x})\right] - \mathbb{E}_{p^*(x)}\left[\ln p_{\mathcal{Z}}(g(\boldsymbol{x}; \boldsymbol{\phi})) + \ln \left| \det \frac{\partial g(\boldsymbol{x}; \boldsymbol{\phi})}{\partial \boldsymbol{x}} \right| \right], \end{aligned} \tag{6.12}$$

The first term should be a constant for a fixed set of samples and thus can be ignored. The second term is the likelihood of the data; it can be evaluated stochastically by Monte Carlo method: drawing samples from $p^*(\boldsymbol{x})$, which is the training data supposed to be, and computing the mean. Optimizing over the forward KLD is therefore equivalent to maximizing the likelihood. It is useful when the target distribution is not explicitly known but can be sampled from, which is the case in GW parameter estimation. The reverse KLD $D_{\mathrm{KL}}\left(p_{\mathcal{X}}(\boldsymbol{x} \mid \boldsymbol{\phi}) \| p^*(\boldsymbol{x})\right)$ can be used when the target distribution can be explicitly evaluated, such as in variational inference [484].

6.3 Data compression

The normalizing flow for parameter estimation should be conditioned on GW data. However, the raw data is too large to be directly processed by neural networks due to the limited GPU memory, and the information is diffuse, necessitating compression into a more compact form. This includes a dense neural network that reduces the dimension of data and some necessary preprocessing of data.

6.3.1 Embedding layers

Singular value decomposition

For 8-second time domain data sampled at 2048Hz, each time series will consist of 16384 data points, which could be a large input for neural networks, even for the embedding network. For example, if we attempt to reduce the data points by half, mapping the 16384 points to 8192, this would require a 16384×8192 matrix, containing approximately 130 million learnable parameters. Allocating such extensive resources to a single linear layer is impractical.

To overcome this problem, Singular Value Decomposition (SVD) is often applied to extract the principal components of the signal. Consider a large number (m) of waveforms $\mathbf{h_i}$, $i = 1, 2, \ldots, m$ in the rows of a $m \times n$ matrix $\mathbf{H}$, where n is the length of each waveform, we can always perform SVD to $\mathbf{H}$:

$$\mathbf{H} = \mathbf{U}\mathbf{S}\mathbf{V}^*, \tag{6.13}$$

where $\mathbf{U}$ and $\mathbf{V}^*$ are complex unitary matrix of shape $m \times m$ and $n \times n$, respectively. The * denotes conjugate transpose. $\mathbf{S}$ is a $m \times n$ rectangular diagonal matrix with non-negative real numbers on the diagonal, which are referred to as *singular values*. Assuming $\mathbf{H}$ has a rank of r, then the first r columns of $\mathbf{V}$ are a set of bases of the row space of $\mathbf{H}$. In other

words, the columns of $\mathbf{V}$, denoted as $\mathbf{v_i}$, can be used as linear bases of waveforms:

$$\mathbf{h_i} = \alpha_{i1}\mathbf{v_1} + \alpha_{i2}\mathbf{v_2} + \dots, \tag{6.14}$$

where the coefficients satisfy

$$\alpha_{ik} = \mathbf{h_i} \cdot \mathbf{v_k} \tag{6.15}$$

because $\{\mathbf{v_k}\}$ are unitary and orthogonal. If we truncate $\{\mathbf{v_k}\}$ to the first l bases (ranked by the corresponding singular values), and use the truncated bases to reconstruct $\mathbf{H}$, the error is [258]

$$\|\mathbf{H} - \mathbf{H_{reconstructed}}\| = \sqrt{\sum_{i=l+1}^{r} (s_i)^2}, \tag{6.16}$$

where s_i is the i'th singular value and the $\|\dots\|$ is the Frobenius norm

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |A_{ij}|^2}. \tag{6.17}$$

This allows us to select $\{\mathbf{v_k}\}$ that corresponds to larger singular values and project the waveform onto the selected bases, and use the projections $\{\alpha_{ik}\}, k = 1, \dots, l$ to reconstruct the original data $\mathbf{h_i}$ with a controllable loss of information. Ref. [357] shows that 100 complex bases are enough to reconstruct 8-second BBH signals in the frequency domain, making the input data suitable for machine learning models.

The SVD projection is linear: it naturally fits the GW data where signals and noises are added linearly. The projections of real data onto SVD bases consist of the projections of signal (Eq. 6.15) and the projections of the noise. For Gaussian noise, its projections, which are linear combinations of Gaussian variables, remain the same Gaussian distribution because the bases are unitary and orthogonal. This nice property is important for efficient data simulation in the training stage, which will be discussed in Sec. 6.4.

Neural dense network

Given that there are multiple sets of projections for a detector network, further data compression is necessary. The SVD projections can be input into a neural network for this purpose. Typically, the embedding network first maps the SVD projections to a higher-dimensional space, from which it extracts features, and then compresses these features into a lower-dimensional space. For example, `DINGO` uses a deep residual MLP as the embedding layer with the first linear transformation initialized as the matrix $\mathbf{V}$, while Ref. [359] introduces an extra CNN. Other viable options, such as transformers[334, 485], are also worth exploring.

The compressed features, usually at the order of hundreds, will serve as the condition for the inference. In this work, we will explore two types of embedding networks: deep residual MLP and Vision Transformer (ViT) [485]. ViT is an extension of the transformer architecture to image processing, but can also be applied to one-dimensional sequence data processing. ViT divides the input data into fixed-size patches. Data is processed by residual MLPs within each patch, and different patches are then embedded into vectors and processed by the transformer encoder. ViT takes advantage of the transformer's ability to capture global relationships inside data and is more memory efficient as data is divided into patches.

6.3.2 Preprocessing extra-long data

For frequency domain complex BBH waveforms, Ref. [357] shows that 100 SVD bases are sufficient to recover the full signal. However, linear SVD bases are less effective for much longer BNS signals due to the highly oscillatory nature of the real and imaginary components compared to BBH waveforms. Besides, the large number of data points makes it challenging to compress the signal into a shorter array. Fig. 6.3 shows the frequency-domain BBH and BNS waveforms from 50Hz and the accuracy of SVD-reconstructed waveforms with different number of bases. While BBH waveform is smooth and can be accurately represented by 32 bases, the BNS waveforms are more complex and require more than 500 bases to accurately reconstruct. As the oscillations get denser at low frequencies (Eq. 1.95), achieving accurate and efficient SVD reconstruction for full-bandwidth waveforms from as low as 20 Hz or 5 Hz becomes impractical. Consequently, the extra-long BNS data in the 3G detectors need to be preprocessed to a reduced size before being put into any neural network. Following Dax et al. [486], we will employ multi-banding and heterodyne prior to generating the SVD bases.

Multi-banding

Multi-banding leverages the observation that CBC signals exhibit low frequencies in the early stages and higher frequencies as they approach the merger, with the signal frequency increasing monotonically over time (assuming quasi-circular orbits). Therefore, coarser frequency resolutions (shorter time durations) can be employed in high-frequency regions to reduce the number of data points.

We have demonstrated a simple multi-banding scheme in Fig. 5.5 in Chapter 5 for rapid pre-merger localization of long BNS signals. More elaborate multi-banding schemes are proposed for parameter estimation in Morisaki [463], Vinciguerra et al. [487]. Specifically, Dax et al. [486] follows Vinciguerra et al. [487], reducing the frequency resolution by a factor of 2 from its neighboring preceder band to achieve a lower frequency resolution

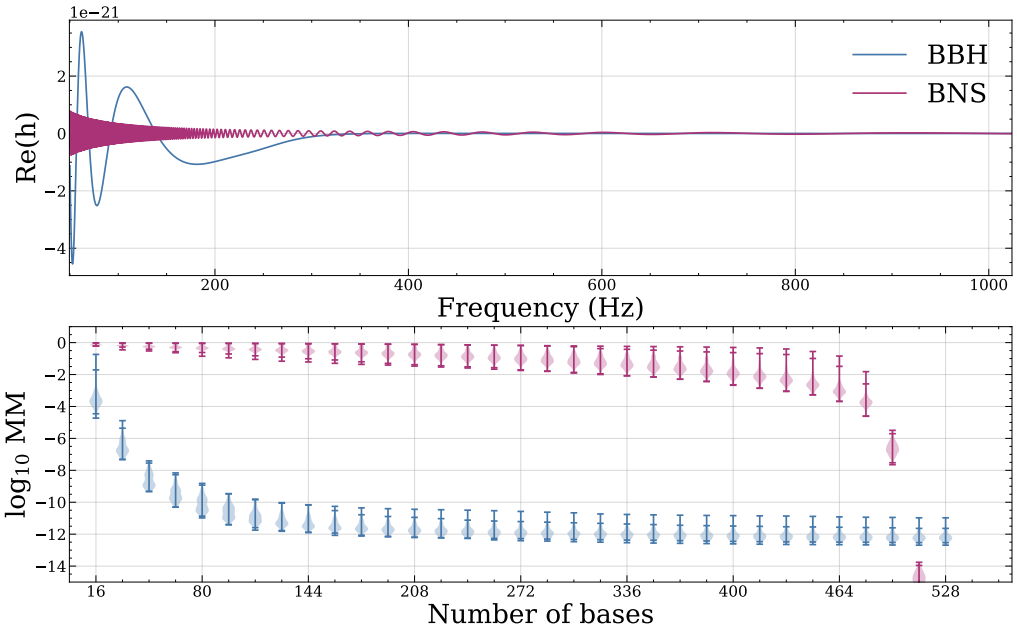


Figure 6.3: Comparisons between frequency-domain waveforms and SVD reconstruction accuracy of BBH and BNS waveforms from 50Hz. **Upper panel**: Real parts of frequency-domain waveforms from 50Hz. **Lower panel**: mismatch of reconstructed waveforms with different number of SVD bases. The SVD bases are calculated with 30000 waveforms generated by `IMRPhenomD` model from 50Hz, assuming zero spin. Black hole mass range is uniformly sampled in $8M_\odot$ and $60M_\odot$, and neutron star mass is between $1M_\odot$ and $2M_\odot$. The BBH waveform contains ∼8000 complex data points while the BNS waveform contains $\sim$ 30000.

in higher frequency regions. However, this method does not fully exploit the frequency evolution of CBC signals as the factor of 2 is a choice for convenience rather than optimization. In this work, we propose a novel multi-banding scheme that adaptively selects frequency nodes and resolutions, ensuring that each band's resolution is precisely tuned to the needs of BNS signals. Detailed implementation of this method is provided below.

Data is evenly sampled within each band, with the number of data points per band set to a customizable constant N. If the parameters of a BNS system are perfectly known, we know the exact frequency evolution of this system and N can be set to a very small value such that the frequency resolution continuously varies between bands, and the data is sampled perfectly aligned to the $\tau(f)$ relation (Eq. 1.82) [487]. However, since we aim for the multi-banding scheme to handle a range of sources with varying masses and spins, and we never perfectly know the source parameters in real detection, N should be set to a larger value. In this work, we choose $N = 64$ to accommodate the unknown source parameters. While a larger N provides more flexibility and robustness, it compromises the efficiency of data compression, and *vice versa*.

Having N set, we search from the highest frequency cut-off f_0, to a lower frequency f_1 such that

$$\alpha_{\text{safety}}(f_0 - f_1)\left[\tau(f_1) - \tau(f_0)\right] = N, \tag{6.18}$$

where $\tau(f)$ is the time-to-merger function. Eq. 1.82 gives the relation to the Newtonian order, but in this chapter, we employ the $\tau(f)$ relation to the 3.5PN order for higher accuracy

$$\tau_{\text{3.5PN}}(f) = t_c - \frac{1}{2\pi}\frac{d\psi_{\text{3.5PN}}(f)}{df} \tag{6.19}$$

where $\psi_{\text{3.5PN}}(f)$ is given by Eq. 1.147. The $\alpha_{\text{safety}} > 1$ is a safety factor that enlarges the effective band duration, ensuring that the frequency resolution is high enough to cope with the potential errors in $\tau_{\text{3.5PN}}(f)$ and the source parameters. We set $\alpha_{\text{safety}} = 2$ in this work. The frequency resolution in the band $(f_1, f_0]$ is given by

$$\Delta f_0 = \frac{1}{\alpha_{\text{safety}}\left[\tau(f_1) - \tau(f_0)\right]}, \tag{6.20}$$

which corresponds to $T_0 = \alpha_{\text{safety}}\left[\tau(f_1) - \tau(f_0)\right]$ data in the time domain. In practice, we should take time-domain data of duration T_0 to generate the frequency-domain data, and make sure the segment contains signals from f_1 to f_0, which is feasible thanks to the safety factor.

This process can be repeated to obtain $f_2,\ f_3,\ \ldots$ and the corresponding Δf_i, until the lower frequency bound reaches the lower frequency cut-off (5Hz). With this scheme and the aforementioned setup, we can compress the original frequency-domain GW waveform, which spans 12 million data points (from 5 Hz to 1024 Hz over a 12000-second duration),

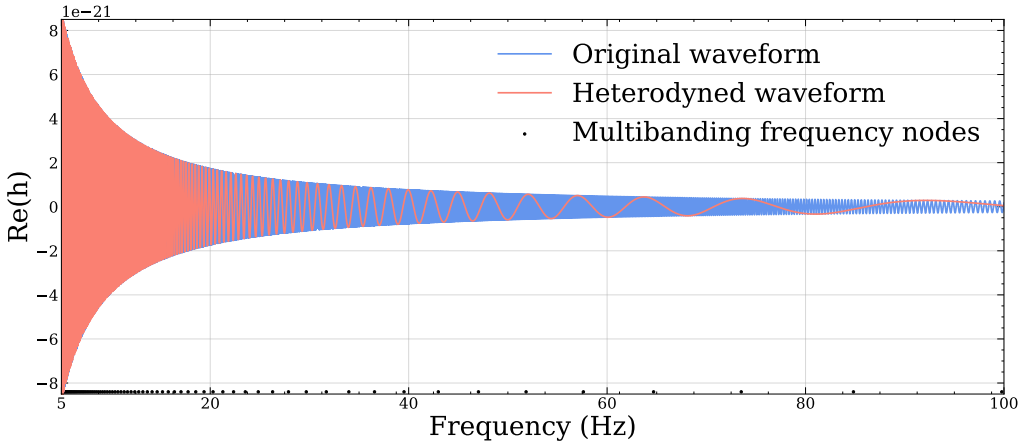


Figure 6.4: BNS frequency-domain waveforms before (blue) and after (orange) heterodyning. Waveforms are truncated at 100Hz for better illustration. Black dots denote the frequency nodes of the multibanding scheme: the bands are defined between nodes.

down to 6000 data points, achieving a compression ratio of 2000. The full-bandwidth GW waveform can be accurately reconstructed by interpolating the amplitude and phase. An illustration of the multibanding frequency nodes is shown in Fig. 6.4. The bands are defined between nodes and data is sampled with frequency resolution in analogy to Eq. 6.20 in each band.

Heterodyning

The multi-band GW waveform is still highly oscillatory. The oscillations in the frequency domain are described by $e^{i\psi(f)}$, where $\psi(f)$ is given by Eq. 1.147 to the 3.5PN, the leading order (0PN) of which is determined by the chirp mass of the system (Eq. 1.95), while the higher orders require the mass ratio. In practice, the chirp mass is the most precisely constrained parameter for BNS systems, typically with relative errors at the sub-percent level or even lower. It can be estimated before performing full parameter estimation, for instance, through matched filtering or a flow-based estimation of the narrow-band data, where a direct SVD is feasible. The pre-estimated chirp mass can be used to simplify the signal by multiplying it with a factor:

$$h_0(f, \mathcal{M}) = e^{i\frac{3}{128}(\frac{\pi G\mathcal{M}f}{c^3})^{-5/3}}, \tag{6.21}$$

which cancels the dominant oscillatory term in the GW waveform, resulting in a slowly varying signal. This technique is essentially *heterodyning* and has been used in the lazy likelihood [488] and relative binning [489, 490]. Fig. 6.4 shows the BNS waveform before and after heterodyning.

As the heterodyning removes some information about the source chirp mass $\mathcal{M}$, the

chirp mass used in heterodyning $\tilde{\mathcal{M}}$ should be provided to the normalizing flow as an extra condition. Given that the pre-estimated $\tilde{\mathcal{M}}$ is not perfectly accurate, the model should be able to deal with a certain level of uncertainty in $\tilde{\mathcal{M}}$. In other words, the model should be trained with $\tilde{\mathcal{M}} \in [\mathcal{M} - \delta\mathcal{M}, \mathcal{M} + \delta\mathcal{M}]$, where $\delta\mathcal{M}$ is a hyperparameter of the model and should be greater than the statistical error bar $\Delta\mathcal{M}$. During inference, one can divide the entire chirp mass space into many intervals $[\mathcal{M}_i - \delta\mathcal{M}, \mathcal{M}_i + \delta\mathcal{M}]$, and perform heterodyning and conditioning with different $\mathcal{M}_i$, then choose the $\mathcal{M}_i$ that results in the highest likelihood in the inference. This technique is referred to as *prior-conditioning* and is described in Ref. [486].

In principle, waveforms can be further smoothened by introducing higher order heterodyning phase terms. However, according to Eq. 1.147, the mass ratio, which is not always well constrained in parameter estimation, is required to compute phases of higher PN orders. Moreover, introducing the mass ratio would result in a search in the two-dimensional space $(\tilde{\mathcal{M}}, \tilde{q})$, bringing additional computation burden. We therefore do not include higher PN orders in heterodyning.

Finally, we note that the heterodyning term $h_0(f, \tilde{\mathcal{M}})$ does not change the statistical property of Gaussian noise as it only induces a phase change, while frequency-domain Gaussian noise has a uniform distribution in phase.

SVD after preprocessing

With multibanding and heterodyning, the 12M (12 million) data points and highly oscillatory GW waveforms are simplified to relatively smooth signals of 6000 data points, on which SVD can be performed. We use the random SVD algorithm [491] implemented in the `scikit-learn` package. For the parameter space $1M_\odot \leq m_{1,2} \leq 3M_\odot$ (detector frame), isotropic spin direction with $0 \leq a_{1,2} \leq 0.05$, $0 < \Lambda_{1,2} \leq 5000$ and isotropic θ_{JN}, SVD can reconstruct the waveform with ~ 300 bases, as shown in Fig. 6.5. Further constraining the mass to $2M_\odot \leq \mathcal{M} \leq 2.1M_\odot$, $0.5 \leq q \leq 1$ and the tidal parameter $\tilde{\Lambda} < 1600$, less than 80 bases would be enough. This means we can effectively compress the extra-long BNS data in the 3G detectors into a manageable size for machine learning models with a combination of multibanding, heterodyning, and SVD techniques.

Note that the SVDs shown in this section are for waveforms only and do not contain extrinsic parameters (luminosity distance is fixed at 1Mpc). Including variations of extrinsic parameters means performing SVD for signals in the detector. This is still manageable but requires $\sim 20\%$ more bases for accurate reconstruction. In this work, we build two sets of SVD bases: one is for waveforms only and is used to compress and decompress the waveforms, and the other is detector-specific and contains variations of $(\alpha, \delta, \psi, t_c)$ but not d_L, and is used to compress the GW signals.

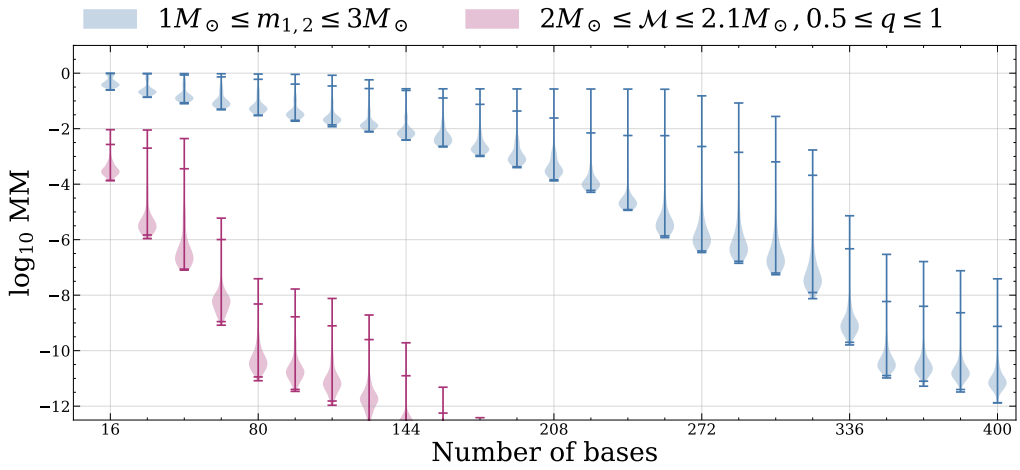


Figure 6.5: Similar to Fig. 6.3, but for the full-bandwidth BNS waveform SVD reconstruction with heterodyning and multibanding. Different colors represent two different mass priors. The narrower mass range needs fewer bases to reconstruct.

6.4 Building the neural network

In this section, I describe the technical details of building the flow-based neural density estimator, including the method of dealing with large amounts of data and training the model.

6.4.1 Dataset

We should provide the model with enough examples of GW data and the corresponding source parameters so that the model can learn data-conditioned transformations that convert the source parameter to the standard Gaussian distribution in the latent space. Given the high dimensionality of GW parameter space, a huge amount of samples are required to cover the entire parameter space and avoid overfitting. Following `DINGO`, we split parameter space into intrinsic and extrinsic parameters and we simulate new data with random extrinsic parameters during training. Details are given below.

Intrinsic parameters

The intrinsic parameters here are defined in the sense that they cannot be easily factored out in the signal, i.e., changing an intrinsic parameter to a new value would induce a non-trivial change in the signal. There are 12 intrinsic parameters in total for BNS systems: component masses $(\mathcal{M}, q)$, six spins (in the system frame, see Sec. 1.4.1), two tidal parameters $(\tilde{\Lambda}, \delta\tilde{\Lambda})$, inclination θ_{JN} and phase ϕ_c. Here we include the spin precession and tidal effects, but ignore other effects like orbital eccentricity and higher modes.

Since generating a large amount of waveforms is time-consuming even with the aid of

multibanding, computing waveforms on-the-fly is impractical. Instead, we draw samples in the intrinsic parameter space and calculate (heterodyned) waveforms in advance, and store the waveforms' projections onto SVD bases to disk. During real-time data simulation, these waveforms can be read from disk and reconstructed by SVD bases, which is fast because SVD reconstruction is a matrix multiplication that can be easily parallelized.

The sampled parameter space is the prior distribution for the neural density estimator, as the estimator only encounters parameters from those regions. The posterior distribution can be seen as the prior-weighted likelihood. Since the prior plays the role of weight, it can be factored into a separate weight (another prior distribution) by applying importance sampling to the samples from the neural density estimator. The shape of the prior distribution is not important in this sense so it can be chosen for the convenience of training the model, but the width of the prior (the size of parameter space) could make a difference. In this chapter, we employ two types of prior:

- **Prior 1**: A large, exploratory prior. $m_1, m_2 \sim \mathrm{U}(1M_\odot, 3M_\odot)$ in the detector frame. $\Lambda_1, \Lambda_2 \sim \mathrm{U}(0, 5000)$. $a_1, a_2 \sim \mathrm{U}(0, 0.05)$ as BNS are expected to have low spins. Here $\mathrm{U}(a, b)$ denotes uniform distribution between a and b. Angles are sampled based on their nature (uniform or isotropic). Masses are converted to $\mathcal{M}$ and q and tidal parameters are converted to $\tilde{\Lambda}$ and $\delta\tilde{\Lambda}$ during training and inference.

- **Prior 2**: A finer prior. $\mathcal{M} \sim \mathrm{U}(2M_\odot, 2.1M_\odot)$ in the detector frame. $q \sim \mathrm{U}(0.5, 1)$. $\tilde{\Lambda} \sim \mathrm{U}(0, 1600)$ and $\delta\tilde{\Lambda}$ is sampled uniformly in the allowed range (conditioned on $m_1, m_2, \tilde{\Lambda}$ and satisfies $\Lambda_{1,2} > 0$). Spins and angles are the same as prior 1.

The prior 1 serves as an exploratory trial to assess the capability of the current network structure. For this, we draw 16 million intrinsic parameter samples for training. However, we discovered that the prior might be too broad to be effectively represented with a manageable number of training samples (see Sec. 6.5), so we proposed prior 2, which is much narrower and we upscale the sample size to 64 million to avoid overfitting. The SVD bases for these two prior distributions are constructed and shown in Sec. 6.3.2.

Extrinsic parameters

There are 5 extrinsic parameters: right ascension α, declination δ, polarization angle ψ, luminosity distance d_L, and coalescence time at geocenter t_c. Once the waveforms are reconstructed, we can readily rescale and project them according to these extrinsic parameters. Therefore, the extrinsic parameters can be generated randomly during training,, effectively allowing for an infinite number of samples in these dimensions.

The prior for (α, δ) is chosen to be isotropic on the sky, and $\psi \sim \mathrm{U}(0, \pi)$. The distance prior needs to be constrained with a certain range to control the SNR, as PE models for high SNR events are particularly challenging to train [486]. However, directly

sampling the luminosity distance can lead to significant variability in observed SNR due to random inclination angles and sky positions, potentially introducing extreme cases such as subthreshold or high-SNR outliers in the training set, which can affect training efficiency. Furthermore, a very large distance prior would require training multiple models to cover the entire detector horizon, complicating inference since the distance is unknown beforehand. Therefore, we do not directly sample the luminosity distance in this work, instead, we set a range for the network SNR and sample the SNR uniformly. With other parameters sampled, the luminosity distance is then rescaled to a value such that the signal would produce the desired network SNR. This controls the signal loudness that the model encounters during training, making the model better adapt to its training domain. The network SNR range is set to be $[20, 50]$ in this work. The majority of relatively informative BNS events will fall into this SNR range, and we leave the high SNR challenge to future works.

The coalescence time is represented by GPS time, which ranges over several billion seconds and varies by 86400 seconds per day, is too large to be directly managed by a single neural network. Following Ref. [492, 493], we make use of geometric symmetry to simplify the problem. The sky location (α, δ) of an astronomical source is considered fixed in the celestial frame, and variations in coalescence time are essentially the rotation of the Earth. Since the detector response functions depend only on the relative position between the detector and the source, the geometry can be equivalently described by fixing the Earth and rotating the source around it. The trace of the source is parallel to the celestial equator, so changes in coalescence time can be mapped to changes in the right ascension of the source. One can select a reference coalescence time and assume the Earth is fixed there, and rotate the right ascension of the source to mimic the changes in time. This is illustrated in Fig. 6.6: the response of the detector at $t_{\rm ref}$ to the source at $(\alpha_{\rm ref}, \delta)$ is the same as the detector at t_c responses to the source at (α, δ). Mathematically:

$$F_{+,\times}(\alpha, \delta, \psi, t_c) = F_{+,\times}(\alpha_{\rm ref}, \delta, \psi, t_{\rm ref}), \tag{6.22}$$

and the difference in time is compensated by changes in right ascension:

$$\alpha - \alpha_{\rm ref} = {\rm GAST}(t_c) - {\rm GAST}(t_{\rm ref}), \tag{6.23}$$

where GAST is the Greenwich Apparent Sidereal Time (GAST), obtained by adding a nutation correction in right ascension to the GMST [494].

For our model, we choose $t_{\rm ref} = 0$ and time prior of ${\rm U}(t_{\rm ref} - 10{\rm ms}, t_{\rm ref} + 10{\rm ms})$. The model will infer the coalescence time within the prior and obtain the corresponding $\alpha_{\rm ref}$, which can be further converted to t_c and real source right ascension α by Eq. 6.23.

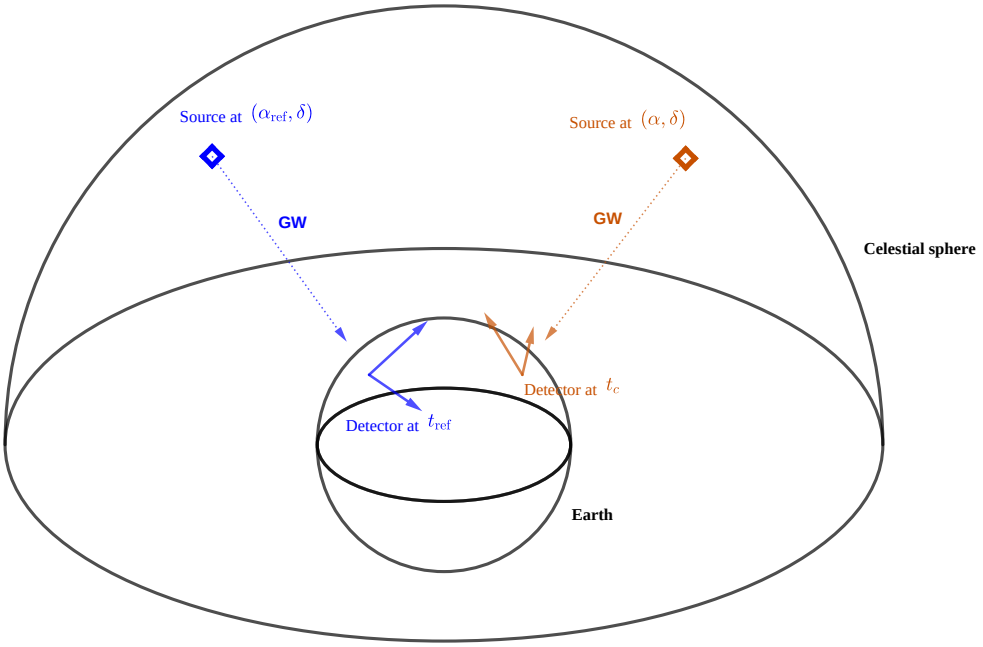


Figure 6.6: Illustrations of reference time. The signal of the detector at $t_{\rm ref}$ seeing the source at $(\alpha_{\rm ref}, \delta)$ is the same as the detector at t_c seeing the source at (α, δ).

Fast data simulation

We calculate waveforms on the intrinsic parameter grids and save their SVD projections to disk. The GW data can then be rapidly simulated during training. We consider the same detector network as Chapter 5, i.e., 1ET+2CE. The process is described as follows.

The GW data d consists of signal h and noise n:

$$d(f) = h(f) + n(f), \tag{6.24}$$

The signal can be calculated by reconstructing the waveform and applying time shifts and detector response. Denote the heterodyned waveform SVD projections as $\boldsymbol{\alpha}_{+,\times}$, i.e.,

$$\boldsymbol{\alpha}_{+,\times} = h_{+,\times}(f, d = 1\text{Mpc}) h_0(f, \mathcal{M}) \cdot \boldsymbol{V}_{\mathbf{wf}}. \tag{6.25}$$

where $h_0(f, \mathcal{M})$ is the heterodyning waveform Eq. 6.21 and $\boldsymbol{V}_{\mathbf{wf}}$ is the SVD bases matrix for waveforms which does not contain variations of extrinsic parameters. We also have detector-specific $\boldsymbol{V}_{\mathbf{det}}$ for signals in detectors and they are trained with varying extrinsic parameters $(\alpha, \delta, \psi, t_c)$. d_L is always fixed at 1Mpc for obtaining these SVD bases. The

signal can be simulated by

$$\begin{aligned} h(f) &= [F_+(\alpha,\delta,t_c,\psi)h_+(f) + F_\times(\alpha,\delta,t_c,\psi)h_\times(f)]\, e^{-2\pi i f[t_c+\Delta t(\alpha,\delta,\tau_{3.5\mathrm{PN}}(f))]} \\ &= \frac{1}{d_L}\left[F_+\boldsymbol{\alpha}_+ \cdot \boldsymbol{V}^*_{\mathbf{wf}}\, h_0^{-1} + F_\times\boldsymbol{\alpha}_\times \cdot \boldsymbol{V}^*_{\mathbf{wf}}\, h_0^{-1}\right] e^{-2\pi i f[t_c+\Delta t]}, \end{aligned} \tag{6.26}$$

where $(\alpha, \delta, \psi, t_c, d_L)$ are randomly sampled during training[1]. Δt is the time delay from geocenter introduced in Eq. 1.120. It is a function of the source sky direction and arrival time, but the latter should be written as a function of frequency $\tau(f)$ due to the long signal duration. The time-frequency relation is given in Eq. 6.19.

We then heterodyne, whiten and project the data. Response functions and time delays are calculated using the scheme described in Sec. 1.3.1. For multi-band strain, whitening should take different noise variances in each band into consideration:

$$h_{\mathrm{whiten}}(f_i) = \frac{h(f_i)}{\sqrt{\frac{T_i}{4}S_n(f_i)}}, \tag{6.27}$$

where f_i is the i'th frequency step and T_i is the corresponding duration of the band f_i belongs to. The heterodyned and whitened signal can then be projected to SVD bases via $\boldsymbol{V}_{\mathbf{det}}$:

$$h_{\mathrm{SVD}} = (h_{\mathrm{whiten}} h_0(\tilde{\mathcal{M}})) \cdot \boldsymbol{V}_{\mathbf{det}}, \tag{6.28}$$

where $\tilde{\mathcal{M}}$ is the chirp mass used for heterodyning. We keep the first 400 (128) projections for prior 1 (2), i.e., h_{SVD} is an array of length of 400 (128). The whitened noise n_{whiten} simply follows a standard Gaussian distribution:

$$n_{\mathrm{whiten}}(f) \sim N(0,1) + iN(0,1), \tag{6.29}$$

and the distribution does not change after heterodyning and SVD projection. Therefore, we can directly simulate standard Gaussian noise on the SVD bases:

$$n_{\mathrm{SVD}} = (n_{\mathrm{whiten}} h_0(\tilde{\mathcal{M}})) \cdot \boldsymbol{V}_{\mathbf{det}} \sim N(0,1) + iN(0,1). \tag{6.30}$$

The heterodyned, whitened, and SVD-projected data can then be assembled:

$$d_{\mathrm{SVD}} = h_{\mathrm{SVD}} + n_{\mathrm{SVD}}. \tag{6.31}$$

The d_{SVD} from different detectors are concatenated and sent to the embedding network.

Almost all calculations involved above are array and matrix manipulations, which are well-suited for parallelization over a batch of events. The only exceptions are the

[1] In fact, we first sample the target SNR and the first 4 parameters and calculate the SNR for source at 1Mpc. Then the signal can be rescaled by $d_L = \mathrm{SNR}_{\mathrm{1Mpc}}/\mathrm{SNR}_{\mathrm{target}}$.

calculation of response functions and time delays. They include the conversion between GPS time and the GMST, which relies on a fundamental C library that cannot be easily parallelized. One option is to use multiple CPUs for the conversion, but here we used a linear fit to approximate the conversion, which allows to perform the computation on GPU. We parallelized all the calculations on GPU and managed to simulate 2048 events (including 5 $d_{\rm SVD}$ streams for a 1ET+2CE network) in $\mathcal{O}(0.1)$s. This enables us to deal with large datasets that contain over 10 million events.

6.4.2 Hyperparameters

The properties of neural networks are determined by specifying hyperparameters, which include the size (depth, width) of models, learning rate, batch size, regularization parameters during training and so forth. The choice of hyperparameters can affect not only the accuracy of the model but also its speed, generalizability, and robustness to new data. Ideally, hyperparameters should be fine-tuned to achieve optimal performance. However, training flow models for full parameter estimation is highly computationally demanding and time-consuming, and we lack the computing resources to experiment with too many configurations. In this section, we describe the setups of models that have completed training and performed well in tests. These models may still benefit from further improvement through additional hyperparameter fine-tuning.

Model settings

The detailed model settings and hyperparameters are shown in Table 6.1. As mentioned in Sec. 6.2, we use coupling spline flows for all models. The flow model, CouplingNSF, is parametrized by the number of transformations performed `n_transforms`, the length of the condition `n_conditional_inputs` (contains $\tilde{\mathcal{M}}$ and (`n_conditional_inputs`-1) elements from the embedding layer outputs), and other parameters that specify the complexity of the rational-quadratic splines: the knots and derivatives defined at nodes of `num_bins` bins in $[-$`tail_bound`, `tail_bound`$]$ and are determined by MLPResNets with `n_neurons` dimensions and `nblocks_per_transform` blocks. We marginalize the coalescence phase by ignoring it during the inference, i.e., the normalizing flow learns the marginalized distribution of the `n_input` $= 16$ remaining parameters.

We keep the first `Nbasis` SVD projections and the reconstruction accuracy can be read out from Fig. 6.5. We try two types of embedding layers: MLPResNet and ViT, as introduced in Sec. 6.3.1. MLPResNet contains `Nblock` MLP blocks, and each block includes `depth_per_block` MLPs connected with ResNet structure. Dimensions of MLP layers are given by `middle_features`. The ViT split the input into patches of size of `patch_size`, and each patch is embedded as a vector of length `dim` and goes through

`depth` transformer blocks. The transformer block consists of a multi-head self-attention layer with `heads` `dim_head`-dimensional attention heads and a `mlp_dim`-dimensional MLP. Empirically, the embedding model should be of a similar size to the normalizing flow in terms of trainable parameters.

For prior 1, we trained a small model Prior1-MLPResNet (35.8 million trainable parameters) and a larger model Prior1-ViT (129.3 million trainable parameters). We use $\tilde{\mathcal{M}} = \mathcal{M} + \delta\tilde{\mathcal{M}}$ to heterodyne the data and condition the normalizing flow, where $\delta\tilde{\mathcal{M}} \sim \mathrm{U}(-\delta\mathcal{M}, \delta\mathcal{M})$. As an exploratory experiment, we set $\delta\mathcal{M} = 0$ for these two models, assuming the chirp mass is perfectly known before inference. Based on the tests results (see the next section), we proposed prior 2 and adjusted the hyperparameters accordingly, resulting in two larger models. We set $\delta\mathcal{M} = 0.0005M_{\odot}$ for prior 2 models and used a larger training set.

Training

We use the Adam optimizer [495] for training and employ L2 regularization. The L2 regularization adds a quadratic term (sum of squares of all trainable parameters) to the loss function to penalize a large number of trainable parameters and large values of trainable parameters, which helps to mitigate overfitting. The learning rate is set to 0.0001 initially and is reduced during training using cosine annealing. We also manually turn down the learning rate by a factor of 0.7 if the validation loss has not decreased for 4 epochs. With these settings, it takes approximately 4 hours to complete one epoch (i.e. process 63M simulated-on-the-fly events) on an NVIDIA A100 GPU (80GB RAM) for prior 2 models. The models were trained for 2-3 weeks[2].

6.5 Model performance

We assess the model performance in this section. We will start with prior 1 models and identify the problems, and explain the motivations of the changes in prior 2. We will first check and compare the posterior samples obtained from MLPResNet-embedded and ViT-embedded models, then compare the precision of our parameter estimation with Fisher matrix and SealGW to examine if the model has achieved the optimal state. We perform the P-P test to check model self-consistency. Finally, we will provide statistics on the computational time required for the prior 2 models.

Ideally, the flow models should be validated against full Bayesian parameter estimation, but to date, the stochastic sampling is prohibitively slow for the 3-hour-long BNS

[2]For efficiency, when experimenting with new hyperparameters, we initialize part of the model (either the embedding or the flow) using pre-trained components with the same structures. Thus, the 2-3 weeks is an estimated value for this hierarchical training.

Model	Prior1-MLPResNet	Prior1-ViT	Prior2-MLPResNet	Prior2-ViT
Embedding model	MLP ResNet	ViT	MLP ResNet	ViT
Embedding layer setup	`Nbasis`=400, `Noutput`=127, `Nblock`=16, `depth per block`=2, `middle_features` = [1024]*4 + [512]*4 + [256]*4 + [128]*4	`Nbasis`=352, `Noutput`=127, `patch_size`=32, `dim`=1024, `depth`=12, `heads`=10, `mlp_dim`=2048, `dim_head`=64	`Nbasis`=128, `Noutput`=127, `Nblock`=30, `depth per block`=2, `middle_features` = [2048]*6 + [1024]*6 + ...+ [128]*6	`Nbasis`=128, `Noutput`=127, `patch_size`=32, `dim`=1024, `depth`=10, `heads`=8, `mlp_dim`=2048, `dim_head`=64
Embedding trainable parameters	16.4M	82.3M	72.4M	63.4M
Flow model	CouplingNSF			
Flow setup	`n_input`=16, `n_transforms`=18, `n conditional inputs`=128, `n_neurons`=256, `nblocks per transform`=5, `num_bins`=8, `tail_bound`=1	`n_input`=16, `n_transforms`=15, `n conditional inputs`=128, `n_neurons`=512, `nblocks per transform`=5, `num_bins`=8, `tail_bound`=1	`n_input`=16, `n_transforms`=15, `n conditional inputs`=128, `n_neurons`=750, `nblocks per transform`=5, `num_bins`=8, `tail_bound`=1	`n_input`=16, `n_transforms`=15, `n conditional inputs`=128, `n_neurons`=750, `nblocks per transform`=5, `num_bins`=8, `tail_bound`=1
Flow trainable parameters	19.4M	47.0M	95.6M	95.6M
Total trainable parameters	35.8M	129.3M	168.0M	159.0M
$\delta\mathcal{M}$	0		$0.0005M_\odot$	
Initial learning rate	0.0001			
Batch size	8192		16384	
Training set size (intrinsic parameters)	16M		63M	
Validation set size	131k			
Annealing	CosineAnnealingLR(T_max=200)			
Optimizer	Adam optimizer with L2 regulation (weight_decay=0.7)			

Table 6.1: Hyperparameters of trained models.

signals, and therefore we have to use the Fisher matrix and SealGW as alternative benchmarks. The 17-dimensional Fisher matrix is known to suffer from numerical instability, often predicting error bars that exceed the prior range for some parameters, such as spins [171, 325]. As a result, we limit our comparison to the parameters for which the Fisher matrix provides physically meaningful result, which leads to the following 7 parameters: $\{\mathcal{M}, \tilde{\Lambda}, d_L, \theta_{JN}, t_c, \alpha, \delta\}$. The Fisher matrix forecast may not be accurate for the 17-dimensional problem as we will observe that the flow model can have a more precise chirp mass estimation, but it provides a measure of the model performance to a certain extent.

6.5.1 Prior 1

Fig. 6.7 shows an example corner plot of two prior 1 models. The source is a $1.6M_\odot$+$1.5M_\odot$ BNS system located at 3000Mpc, producing a network SNR of 32. The two models are consistent on most parameters. Both of them are able to put constraints on the masses, spin tilt angles, tidal parameters, and sky locations, and they successfully capture some correlations such as the distance-inclination degeneracy. However, these posterior distributions exhibit several problems.

First, the Prior1-MLPResNet model showed a wider statistical error on the chirp mass, indicating the estimation is not sufficiently precise. The Prior1-ViT model, being larger, performs better in estimating the chirp mass but still falls short of optimal performance. This issue extends to other parameters as well. We simulate 100 events and compare the statistical uncertainties of posteriors given by the flow models with Fisher matrix prediction on $\{\mathcal{M}, \tilde{\Lambda}, d_L, \theta_{JN}, t_c, \alpha, \delta\}$, and compare the skymaps with SealGW. The results are shown in Fig. 6.8. The flow models have larger statistical uncertainties than the predicted values and SealGW calculation. Increasing the model size helps alleviate some of these issues as we notice that the lager model Prior1-ViT gives more precise estimation. However, we find further increases in model size led to overfitting, despite the training set containing 16 million samples with different intrinsic parameters.

Second, both models tend to constrain $\tilde{\Lambda}$ to the interval $[1000, 2000]$, despite the actual value in this example being $\tilde{\Lambda} = 512$. This is because the prior distribution of $\tilde{\Lambda}$ peaks around 2000, and the non-optimal models will lean to the prior distribution because they are not "smart" enough to fully exploit the data (condition) to construct the likelihood.

Third, we assumed $\delta\mathcal{M} = 0$ in prior 1 models, i.e., the data is heterodyned with, and the flow is conditioned on the perfectly known chirp mass. This is not a practical assumption in real detection. In addition, it makes the model heavily rely on the given chirp mass $\tilde{\mathcal{M}}$ and ignore the data, which causes bias in the self-consistency test. We show the P-P plots of two models in Fig. 6.9, and we find the estimation on the chirp mass is typically biased. For Prior1-MLPResNet, the P-P line is steep around 0.5 C.I.,

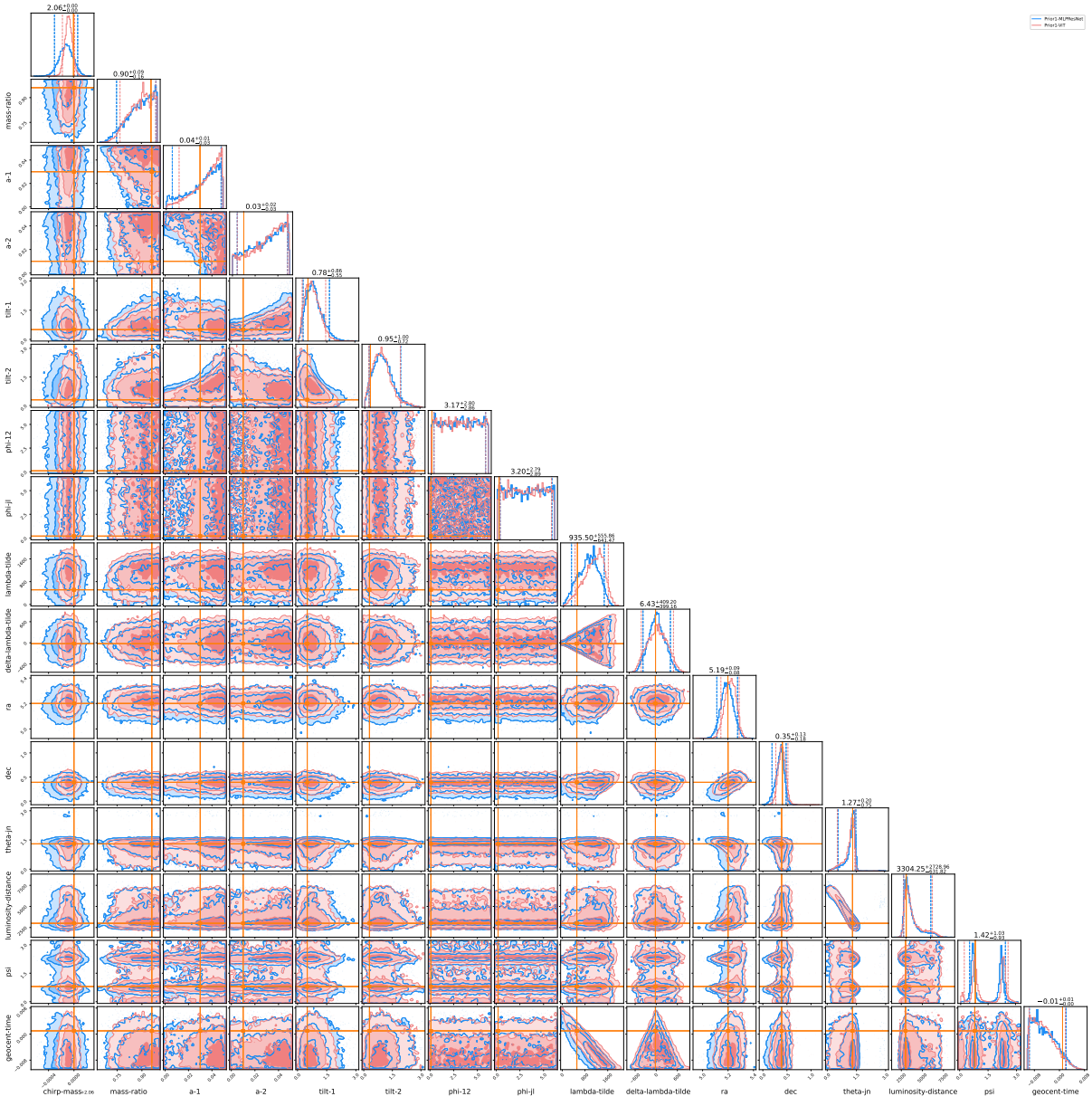


Figure 6.7: An example corner plot of prior 1 models. The blue region is the posterior distribution generated by Prior1-MLPResNet and the red is from Prior1-ViT. The source is a $1.6M_{\odot}$+$1.5M_{\odot}$ BNS system located at 3000Mpc with a network SNR of 32. Dashed lines show the 5% and 95% percentiles. The chirp mass shown is in the detector frame.

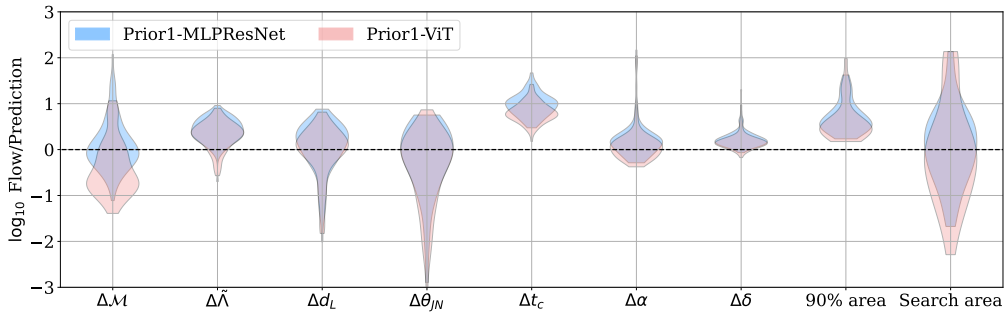


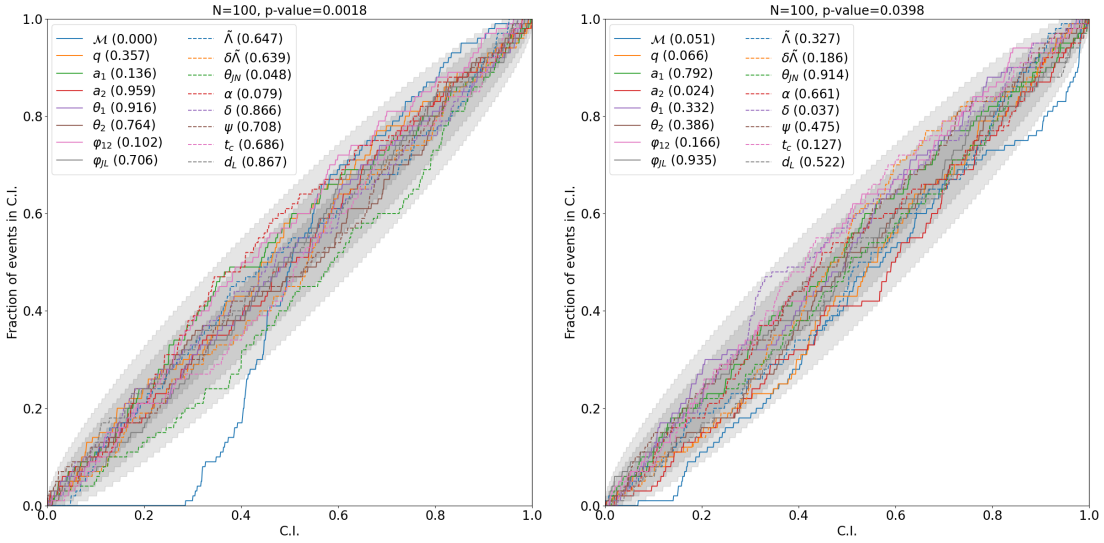
Figure 6.8: Comparing prior 1 models with Fisher matrix and SealGW. The first seven columns display the ratios of the statistical uncertainties in the posterior distributions of the parameters $\{\mathcal{M}, \tilde{\Lambda}, d_L, \theta_{JN}, t_c, \alpha, \delta\}$ provided by the flow models relative to the Fisher matrix predictions. The last two columns compare the skymaps from the flow models with those from SealGW, showing the ratios of the 90% confidence area and searched area (as defined in Sec. 5.2.4). Prior1-MLPResNet is represented in blue, and Prior2-ViT is shown in red. A black dashed line indicates where the flow models' results match those of the Fisher matrix or SealGW. Note that $\Delta\tilde{\Lambda}$ refers to the uncertainty of $\tilde{\Lambda}$, not the source parameter delta lambda tilde ($\delta\tilde{\Lambda}$).

which means the true chirp mass tend to lie in the center of the posterior distribution, suggesting that the model is predominantly reflecting the given chirp mass rather than effectively interpreting the data. Prior1-ViT also shows a systematic bias in its chirp mass estimates.

6.5.2 Prior 2

Based on the observations of prior 1 results, we conclude that 1), we need to enlarge the training set to avoid overfitting and 2), we need to increase the complexity of models or reduce the complexity of training parameter space to fully exploit the data. These insights led to the adjustments made in prior 2, including using narrower chirp mass and tidal parameters ranges, non zero $\delta\mathcal{M}$, larger training set and larger models. Since the chirp mass can be pre-estimated, we could train different models for various chirp mass ranges and apply them based on the pre-estimation. The tidal parameters range is chosen to satisfy the most equations of state of BNS $\tilde{\Lambda} < 1600$ [136], so it is safe as long as the binaries do not show exotic properties.

Example corner plots for the same event from prior 2 models are shown in Fig. 6.10. The two models demonstrate good agreement on this event except for a slight shift in chirp mass distribution. Comparison with Fisher matrix and SealGW is given in Fig. 6.11, which shows a significant improvement on precision of the flow-based models: the logarithm of ratios of statistical uncertainties mostly peak at 0, meaning the flow models give identical results as prediction. Precision on chirp mass surpass the Fisher matrix prediction, im-



(a) Prior1-MLPResNet (b) Prior1-ViT

Figure 6.9: P-P plots of prior 1 models. x-axis is the credible interval and y-axis the fraction of events included in the corresponding credible interval. Digits in the brackets are the p-values and the confidence regions are of 68%, 95%, 99.7% confidence levels. The p-value and confidence region for P-P test are introduced in Sec. 2.2.2.

plying the latter may not be accurate for this 17-dimensional problem. The coalescence time is not optimally constrained. However, the localization is consistent with SealGW, indicating the timing precision is not significantly off. From the corner plot, we notice that it is anti-correlated with tidal parameter $\tilde{\Lambda}$, which is reasonable because systems with a larger tidal parameter will deform and stop the chirp GW emission earlier, leading to an earlier coalescence time. This degeneracy may hinder the model's ability to precisely constrain the coalescence time. Additionally, the flow model produces smaller searched area than SealGW , which means it provides more accurate sky localization for BNS systems. The P-P plot is given in Fig. 6.12. The chirp mass issue is resolved after introducing a non-zero $\delta\mathcal{M}$.

We test the sampling time of the prior 2 models on an NVIDIA RTX 3080 GPU. It takes $\lesssim$ 0.3s to sample 5000 samples, as illustrated in Fig. 6.13. This is a significant improvement over traditional stochastic sampling methods which are still prohibitively slow on long BNS signals to date. The rapid inference capability of these models enables the processing of large volumes of detections in the 3G era with minimal hardware requirements.

The MLPResNet-embedded model and ViT-embedded model exhibit negligible difference in our tests. This is likely due to the fact that the input to the embedding model consists of SVD projections, where the positions of the elements are not crucial. Consequently, the transformer's strength in exploring correlations within a sequence becomes less relevant to this particular problem. However, due to computational constraints, we did not conduct a comprehensive investigation into hyperparameter fine-tuning. The ViT

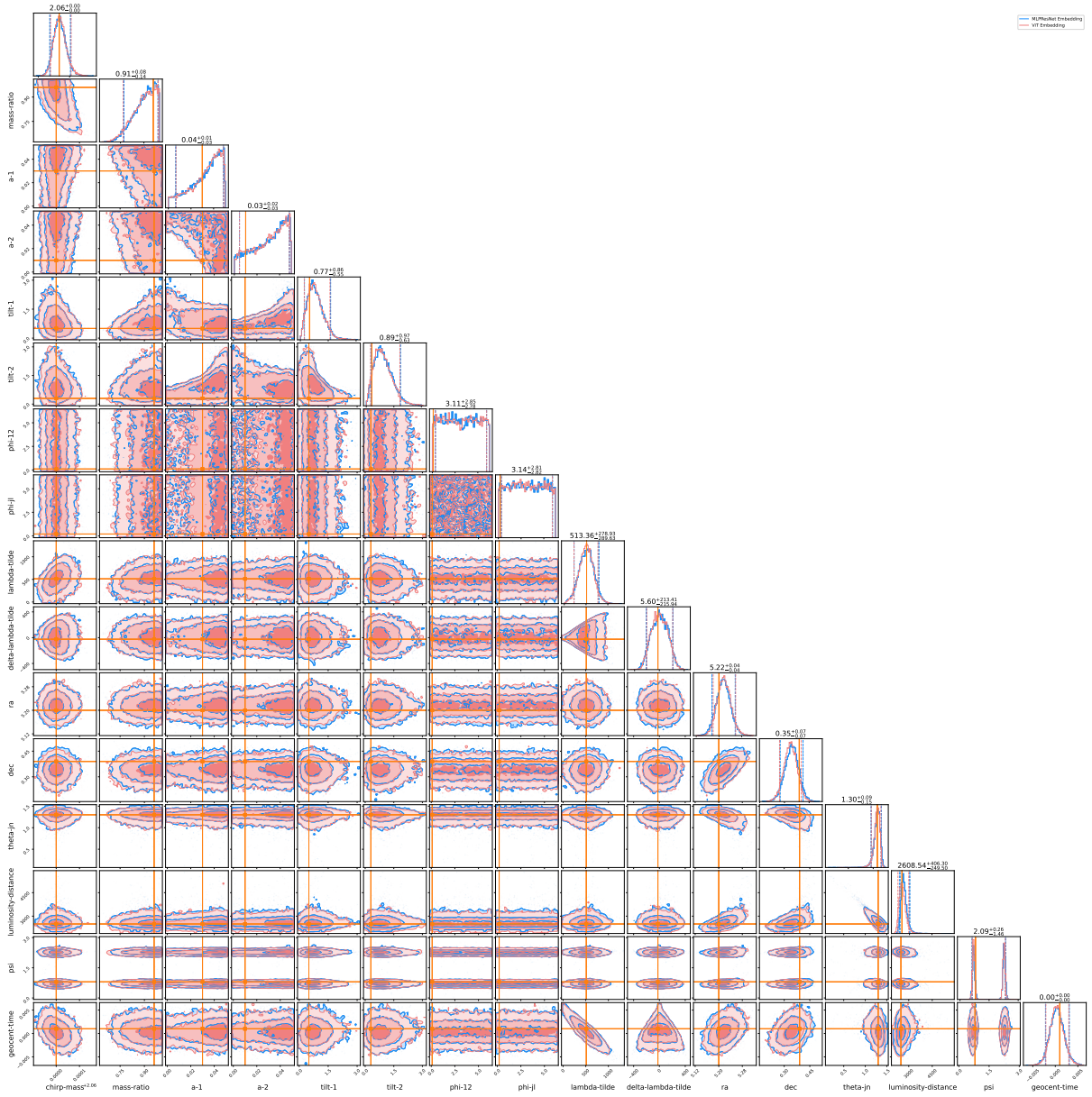


Figure 6.10: Same as Fig. 6.7, but for prior 2 models. The event is the same as Fig. 6.7, but data is heterodyned with $\delta\mathcal{M} = 0.00028 M_\odot$ instead of $\delta\mathcal{M} = 0$.

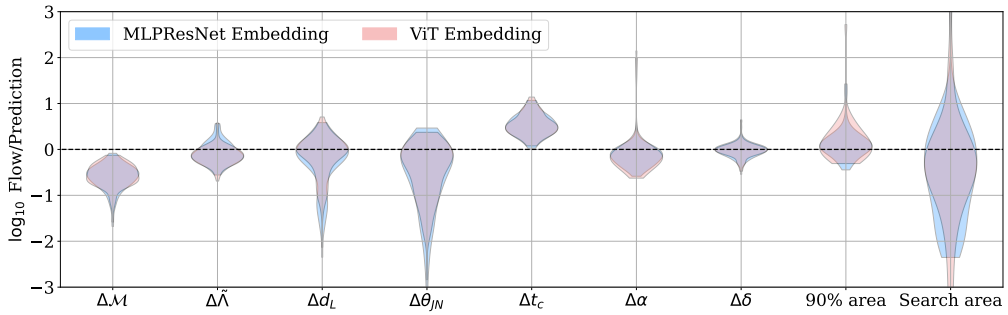
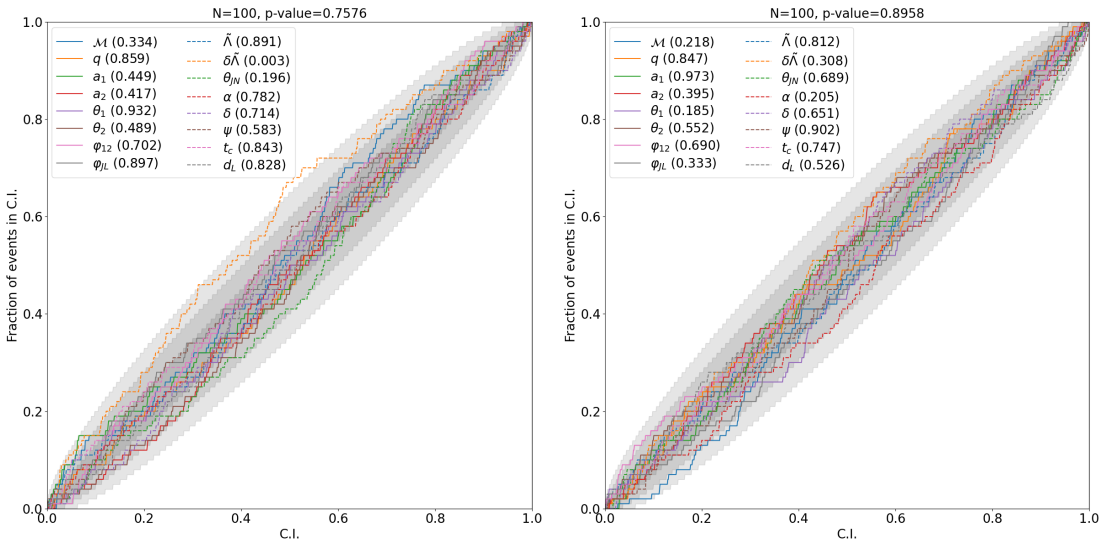


Figure 6.11: Same as Fig. 6.8, but for prior 2 models. The consistency is significantly improved compared with prior 1 models. Most parameters are consistent with theoretical predictions except for chirp mass and coalescence time. Given the models pass the P-P test, we can conclude that the Fisher matrix is not accurate for the high-dimensional problem.



(a) Prior2-MLPResNet (b) Prior2-ViT

Figure 6.12: Same as Fig. 6.9, but for prior 2 models. Prior 2 models have significantly improved self-consistency.

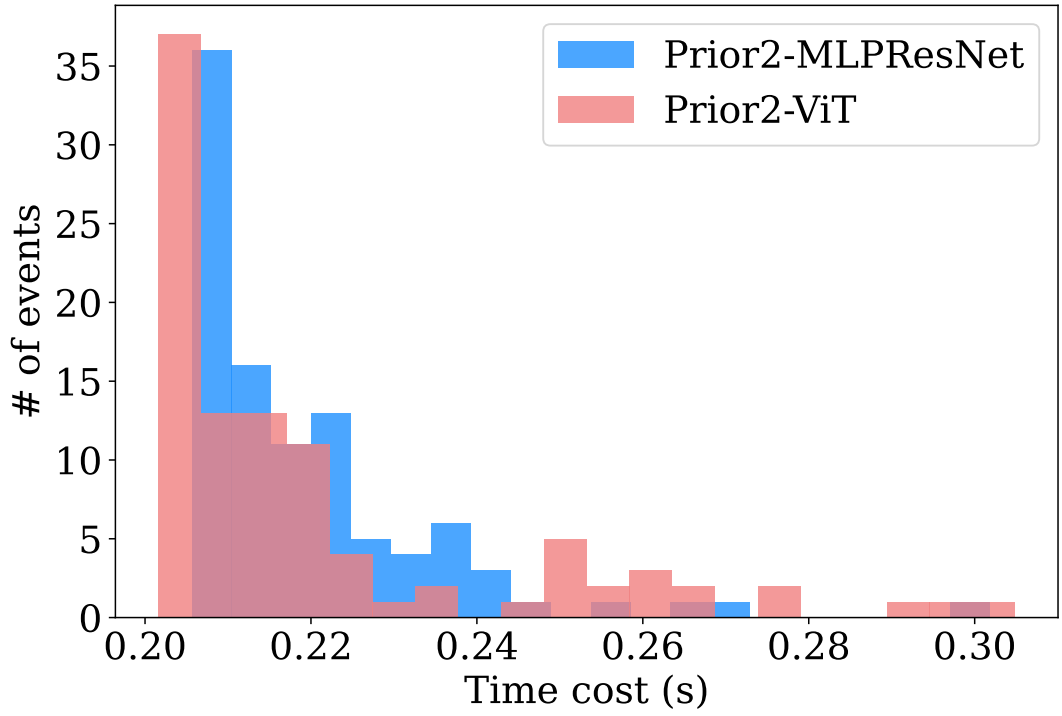


Figure 6.13: Time costs of sampling 5000 samples of prior 2 models. Tests are run on an NVIDIA RTX 3080 GPU.

might demonstrate better performance with larger model sizes, given its architectural advantages. It is also worth noting that the ViT requires significantly more memory and is slower to train compared to MLPResNet.

6.6 Summary and discussions

In this chapter, we demonstrated an exploratory formalism of full parameter estimation of extra-long BNS signals in the 3G detector network (1ET+2CE) using neural density estimation. We utilized the generic network structure that consists of an embedding layer and a normalizing flow for this purpose, and employed multiple techniques to pre-process the long data, including multi-banding with a greedy banding scheme, heterodyning with prior conditioning, and singular value decomposition. We experimented on two types of embedding network (MLPResNet and ViT), but we did not manage to tell the difference in performance between them due to the limited computing resources and the lack of experiments thereof.

We built models for parameter space at different scales: the full parameter space (prior 1) and a constrained parameter space (prior 2) that only spans a small interval of chirp mass. We find an extremely large dataset (exceeding 6×10^7 intrinsic parameters)

is required to prevent overfitting and rapid data simulation is crucial to training the model in a manageable time. Based on our experiments, we concluded that models with current network structure struggle with learning the entire parameter space (prior 1) but they performed well in the constrained one (prior 2), in the sense of agreement with Fisher matrix prediction and SealGW calculation, self-consistency, and sampling time. Specifically, our machine learning models are able to generate 5000 posterior samples within 0.3s. This sheds light on the full parameter estimation problem of extra-long signals in the future detectors, which is exceedingly slow using traditional stochastic sampling methods. By processing large amount of BNS data rapidly, our method has the potential to build catalog-level analysis pipelines for the next generation detectors, and to incorporate with other physics-related analyses, such as constraining the equation of state of neutron stars [361]. In the short term, the method can be applied to the ET and CE Mock Data Challenges and help the design study of 3G detectors.

As an exploratory work, there are several issues that need to be addressed. First, we are not able to compare with the flow-generated posterior with the ground truth generated by stochastic sampling because of the latter's prohibitive computational cost. We made use of approximate alternatives, Fisher matrix and SealGW instead. While these comparisons suggest that our models achieve good precision and retain accuracy, we cannot confirm if they reach the optimal performance or quantify the distance to the ground truth. Sampling methods are actively evolving, and we anticipate that future developments will enable comparisons with long signals in the coming years. This will allow us to better assess and refine our models.

Second, we trained our model in a constrained parameter space: the detector frame chirp mass is within 2 and 2.1 solar masses, and the network SNR is constrained with 20 and 50. As these two quantities are known before parameter estimation, we can, in principle, train multiple networks for different signals. However, the prior range could be further optimized to reduce the total number of required models. In addition, parameter estimation for high SNR events presents additional challenges [486], which will require further investigation. We plan to address these issues in our future work.

Lastly, practical implementation requires addressing additional complexities, such as overlapping signals and noise variations over extended signal durations. We need to investigate how the current model performs under these conditions and whether its structure can effectively handle these challenges. These aspects will be explored in our future work.

The python implementation of the methods described in this chapter can be found in the GitHub repository `river`[3].

[3] https://github.com/MarinerQ/river

Chapter 7

Summary and outlook

In this thesis, I explored robustness and efficiency challenges in the CBC data analysis in the context of high-precision GW astronomy, which is characterized by high SNR and elevated detection rates. My work attempts to address some specific questions within this broad framework. Chapters 1-2 provide an overview of key concepts in GW astrophysics and data analysis and Chapters 3-6 present four distinct projects that contribute to the overall theme. Although each chapter includes its own summary and discussion, I would like to emphasize the most important takeaways from each project here.

Chapters 3 and 4 focus on the systematic errors in GW data analysis. Chapter 3 investigates the false deviation of GR due to inaccurate waveform models and overlapping signals. Our findings are:

- **Error accumulation**: False deviation from GR can occur when results from multiple events are combined (Fig. 3.4).
- **Waveform matters**: Waveform inaccuracy is the primary contributor to systematic errors in the catalog. However, in rare cases where the coalescence times of signals are very close, overlapping signals can also become the dominant source of error (Fig. 3.2, Fig. 3.3).
- **Frequency of signal overlapping**: Signal overlapping rate is estimated (Tab. 3.1), suggesting most signals are not significantly influenced by other overlapping signals. A high overlapping rate would exacerbate the waveform systematics due to incorrect signal subtractions (Fig. 3.6).
- **Golden events**: High-SNR events are particularly susceptible to systematic errors (Fig. 3.3, Fig. 3.5) because the statistical error decreases with increasing SNR while systematics keep constant (Eq. 2.51, Eq. 2.52).

Identifying the importance of waveform accuracy, Chapter 4 introduces a waveform accuracy assessment approach that does not require NR simulations and can be performed in

the entire parameter space. Key ideas and conclusions are:

- **Cross check**: The accuracy of waveform models can be evaluated by comparing different waveform models. A significant difference between models indicates that at least one of the models lacks sufficient accuracy (Fig. 4.1).
- **Current status**: Applying this method to the parameter estimation results of the GWTC-2.1 and GWTC-3 catalogs, we find that the current waveforms could fail our accuracy assessment criterion in certain cases (Fig. 4.2, Fig. 4.4). This method is also being used in O4 results.
- **Impacts on PE**: We found a correlation between our quantified accuracy assessments and systematic errors in parameter estimation (Fig. 4.3).
- **Problematic parameter regions**: Binary systems that exhibit high spin, precession, strong higher modes, and unequal masses are poorly modeled (Fig. 4.4, Fig. 4.5). BNS and NSBH waveforms examined do not show these issues due to the limited inclusion of these complex effects in their waveform modeling.
- **Future requirements**: To meet the demands of future high-precision GW observations, waveform models need to be improved by at least three orders of magnitude in accuracy (Fig. 4.5).

These two chapters emphasize the importance of managing the error budget in the high-precision era of GW astronomy, and showcase the consequences when the errors are not properly controlled. This issue extends beyond tests of GR and affects all aspects where parameter estimation is required. The waveform accuracy assessment method acts as a post-PE check, flagging potential inaccuracies in waveforms based on the result. Although it does not directly solve the waveform issue during PE, it provides a valuable, quantified indication of potential waveform systematics before scientific conclusions are drawn.

In the long term, building highly accurate waveform models would be paramount to controlling systematic errors. Other than that, statistical methods, such as including the uncertainties in waveforms during PE and multi-waveform analysis, would also help mitigate the waveform systematics. Other sources of systematic errors must be carefully managed as well, and a comprehensive review of these challenges is available in Ref. [403].

Chapters 5 and 6 focus on building efficient data analysis methods for long BNS signals. Chapter 5 leverages the SealGW algorithm and a multibanding scheme to achieve pre-merger source localization. The key points from this chapter are:

- **Rotating antenna**: The Earth rotation effects need to be accounted for when analyzing long-duration GW signals (Fig. 5.4).

- **Multibanding and multiplying**: To mitigate issues in matched filtering caused by Earth rotation, we perform matched filtering in multiple bands, within which the Earth rotation can be neglected. Likelihoods derived from the SNR timeseries of each band are then multiplied to improve localization accuracy (Fig. 5.5).

- **Localization statistics**: The event rate and area evolution of early warning localization are estimated for the 1ET+2CE network (Fig. 5.6, Fig. 5.7). Notably, it is possible to provide accurate sky localizations more than 30 minutes before the merger.

- **Fast and cheap**: Our localization algorithm SealGW offers a faster and more resource-efficient approach to processing large numbers of events, making it well-suited for the high event rates expected in future GW observations (Fig. 5.9).

In Chapter 6, we advance to the full parameter estimation for long BNS signals, an extremely computationally expensive task for traditional stochastic sampling methods. We make use of machine learning algorithms to achieve this goal:

- **General structure**: Our model consists of an embedding layer that compresses the data and a normalizing flow that serves as the neural density estimator, sampling from the target distributions (Fig. 6.1).

- **Data pre-processing**: An elaborate multi-banding scheme, heterodyning, and SVD are employed to reduce the size of the data (Fig. 6.4, Fig. 6.5).

- **Big data**: An exceptionally large dataset is required to avoid overfitting and underfitting, necessitating a fast data simulation scheme and the use of high-performance GPU resources.

- **Performance metrics**: Within a constrained parameter space, our trained models are able to generate precise (Fig. 6.11) and accurate (Fig. 6.10, Fig. 6.12) parameter estimates within $\mathcal{O}(0.1)$s (Fig. 6.13).

These works demonstrate the feasibility of analyzing complex data with minimal time and hardware requirements, significantly outperforming traditional methods in terms of efficiency while maintaining a satisfactory level of accuracy. As we approach the 3G era, where we anticipate a large number of detections and high SNRs, efficiency will become increasingly critical in the development of analysis algorithms.

The emergence of novel "non-standard" data analysis methods, which leverage the unique properties of chirp-like signals, (semi-)analytical marginalization, Gaussian approximations, and machine learning algorithms, represents an active area of research in GW astronomy. These methods, offering significant advantages in speed and resource efficiency, have the potential to redefine what is considered "standard" in the future.

However, the "standard" parameter estimation methods, which are the most robust approaches, remain essential. These traditional methods are continually being refined to improve speed and reduce hardware demands. They serve as the benchmark for testing new algorithms and are indispensable for tasks where robustness is paramount, such as in the tests of GR.

Looking ahead, while achieving faster computational speeds is crucial, there is also a pressing need to adapt these algorithms to more realistic scenarios expected in the 3G detectors. This includes addressing challenges like overlapping signals and variations in noise, ensuring that the analysis methods remain robust and reliable under all conditions.

Overall, in this thesis, we addressed the difficulties in the data analysis in high-precision GW astronomy and proposed a range of solutions. As this field is rapidly evolving, we hope that these works provide valuable insights for the ongoing development of data analysis techniques and offer a general picture of key qualities that data analysis methods should possess in the high-precision era.

Bibliography

[1] Qian Hu and John Veitch. Accumulating Errors in Tests of General Relativity with Gravitational Waves: Overlapping Signals and Inaccurate Waveforms. *Astrophys. J.*, 945(2):103, 2023. doi:10.3847/1538-4357/acbc18. (cit. on pp. xii and 131).

[2] Qian Hu and John Veitch. Assessing the model waveform accuracy of gravitational waves. *Phys. Rev. D*, 106(4):044042, 2022. doi:10.1103/PhysRevD.106.044042. (cit. on pp. xii, 78, and 82).

[3] Qian Hu and John Veitch. Rapid Premerger Localization of Binary Neutron Stars in Third-generation Gravitational-wave Detectors. *Astrophys. J. Lett.*, 958(2):L43, 2023. doi:10.3847/2041-8213/ad0ed4. (cit. on p. xii).

[4] Qian Hu, Cong Zhou, Jhao-Hong Peng, Linqing Wen, Qi Chu, and Manoj Kovalam. Semianalytical approach for sky localization of gravitational waves. *Phys. Rev. D*, 104(10):104008, 2021. doi:10.1103/PhysRevD.104.104008. (cit. on pp. xii, 118, 126, and 127).

[5] Charles W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation*. W. H. Freeman, San Francisco, 1973. ISBN 978-0-7167-0344-0, 978-0-691-17779-3. (cit. on pp. xiii, 5, 12, and 18).

[6] Michele Maggiore. *Gravitational Waves. Vol. 1: Theory and Experiments*. Oxford University Press, 2007. ISBN 978-0-19-171766-6, 978-0-19-852074-0. doi:10.1093/acprof:oso/9780198570745.001.0001. (cit. on pp. 2, 5, 13, 18, 20, 24, 26, 43, and 47).

[7] B. P. Abbott et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116(6):061102, 2016. doi:10.1103/PhysRevLett.116.061102. (cit. on pp. 11, 22, 27, 39, and 62).

[8] R. A. Hulse and J. H. Taylor. Discovery of a pulsar in a binary system. *Astrophys. J. Lett.*, 195:L51–L53, 1975. doi:10.1086/181708. (cit. on pp. 15 and 26).

[9] Krzysztof Belczynski, Vassiliki Kalogera, and Tomasz Bulik. A Comprehensive study of binary compact objects as gravitational wave sources: Evolutionary channels, rates, and physical properties. *Astrophys. J.*, 572:407–431, 2001. doi:10.1086/340304. (cit. on pp. 15 and 22).

[10] Joel E. Tohline. The origin of binary stars. *Ann. Rev. Astron. Astrophys.*, 40:349–385, 2002. doi:10.1146/annurev.astro.40.060401.093810. (cit. on pp. 15 and 22).

[11] Clifford M. Will. The Confrontation between General Relativity and Experiment. *Living Rev. Rel.*, 17:4, 2014. doi:10.12942/lrr-2014-4. (cit. on p. 20).

[12] Stephon Alexander and Nicolas Yunes. Chern-Simons Modified General Relativity. *Phys. Rept.*, 480:1–55, 2009. doi:10.1016/j.physrep.2009.07.002. (cit. on p. 21).

[13] Benoit Famaey and Stacy McGaugh. Modified Newtonian Dynamics (MOND): Observational Phenomenology and Relativistic Extensions. *Living Rev. Rel.*, 15:10, 2012. doi:10.12942/lrr-2012-10. (cit. on p. 21).

[14] S. Shankaranarayanan and Joseph P. Johnson. Modified theories of gravity: Why, how and what? *Gen. Rel. Grav.*, 54(5):44, 2022. doi:10.1007/s10714-022-02927-2. (cit. on p. 21).

[15] Antonio De Felice and Shinji Tsujikawa. f(R) theories. *Living Rev. Rel.*, 13:3, 2010. doi:10.12942/lrr-2010-3. (cit. on p. 21).

[16] Claudia de Rham. Massive Gravity. *Living Rev. Rel.*, 17:7, 2014. doi:10.12942/lrr-2014-7. (cit. on p. 21).

[17] Saeed Mirshekari, Nicolas Yunes, and Clifford M. Will. Constraining Generic Lorentz Violation and the Speed of the Graviton with Gravitational Waves. *Phys. Rev. D*, 85:024041, 2012. doi:10.1103/PhysRevD.85.024041. (cit. on p. 21).

[18] Maria Okounkova, Will M. Farr, Maximiliano Isi, and Leo C. Stein. Constraining gravitational wave amplitude birefringence and Chern-Simons gravity with GWTC-2. *Phys. Rev. D*, 106(4):044067, 2022. doi:10.1103/PhysRevD.106.044067. (cit. on p. 21).

[19] Enis Belgacem, Yves Dirian, Stefano Foffa, and Michele Maggiore. Gravitational-wave luminosity distance in modified gravity theories. *Phys. Rev. D*, 97(10):104066, 2018. doi:10.1103/PhysRevD.97.104066. (cit. on p. 21).

[20] Atsushi Nishizawa. Generalized framework for testing gravity with gravitational-wave propagation. I. Formulation. *Phys. Rev. D*, 97(10):104037, 2018. doi:10.1103/PhysRevD.97.104037. (cit. on p. 21).

[21] D. M. Eardley, D. L. Lee, A. P. Lightman, R. V. Wagoner, and C. M. Will. Gravitational-wave observations as a tool for testing relativistic gravity. *Phys. Rev. Lett.*, 30:884–886, 1973. doi:10.1103/PhysRevLett.30.884. (cit. on p. 21).

[22] A. Coley, R. Milson, Vojtech Pravda, and A. Pravdova. Classification of the Weyl tensor in higher dimensions. *Class. Quant. Grav.*, 21:L35–L42, 2004. doi:10.1088/0264-9381/21/7/L01. (cit. on p. 21).

[23] L. D. Landau and E. M. Lifschits. *The Classical Theory of Fields*, volume Volume 2 of *Course of Theoretical Physics*. Pergamon Press, Oxford, 1975. ISBN 978-0-08-018176-9. (cit. on p. 21).

[24] Atsushi Nishizawa, Atsushi Taruya, Kazuhiro Hayama, Seiji Kawamura, and Masaaki Sakagami. Probing non-tensorial polarizations of stochastic gravitational-wave backgrounds with ground-based laser interferometers. *Phys. Rev. D*, 79:082002, 2009. doi:10.1103/PhysRevD.79.082002. (cit. on p. 21).

[25] R. Abbott et al. Tests of General Relativity with GWTC-3. 12 2021. doi:10.48550/arXiv.2112.06861. (cit. on pp. 21, 40, 78, 80, and 91).

[26] Jarrod R. Hurley, Christopher A. Tout, and Onno R. Pols. Evolution of binary stars and the effect of tides on binary populations. *Mon. Not. Roy. Astron. Soc.*, 329:897, 2002. doi:10.1046/j.1365-8711.2002.05038.x. (cit. on p. 22).

[27] Michal Dominik, Krzysztof Belczynski, Christopher Fryer, Daniel Holz, Emanuele Berti, Tomasz Bulik, Ilya Mandel, and Richard O'Shaughnessy. Double Compact Objects I: The Significance of the Common Envelope on Merger Rates. *Astrophys. J.*, 759:52, 2012. doi:10.1088/0004-637X/759/1/52. (cit. on p. 22).

[28] T. M. Tauris et al. Formation of Double Neutron Star Systems. *Astrophys. J.*, 846 (2):170, 2017. doi:10.3847/1538-4357/aa7e89. (cit. on p. 22).

[29] Carl L. Rodriguez, Pau Amaro-Seoane, Sourav Chatterjee, Kyle Kremer, Frederic A. Rasio, Johan Samsing, Claire S. Ye, and Michael Zevin. Post-Newtonian Dynamics in Dense Star Clusters: Formation, Masses, and Merger Rates of Highly-Eccentric Black Hole Binaries. *Phys. Rev. D*, 98(12):123005, 2018. doi:10.1103/PhysRevD.98.123005. (cit. on p. 22).

[30] Hiromichi Tagawa, Zoltan Haiman, and Bence Kocsis. Formation and Evolution of Compact Object Binaries in AGN Disks. *Astrophys. J.*, 898(1):25, 2020. doi:10.3847/1538-4357/ab9b8c. (cit. on p. 22).

[31] Johan Samsing, Morgan MacLeod, and Enrico Ramirez-Ruiz. The Formation of Eccentric Compact Binary Inspirals and the Role of Gravitational Wave Emission in Binary-Single Stellar Encounters. *Astrophys. J.*, 784:71, 2014. doi:10.1088/0004-637X/784/1/71. (cit. on p. 22).

[32] S. J. Kovacs and K. S. Thorne. The Generation of Gravitational Waves. 4. Bremsstrahlung. *Astrophys. J.*, 224:62–85, 1978. doi:10.1086/156350. (cit. on p. 22).

[33] Juan García-Bellido and Savvas Nesseris. Gravitational wave energy emission and detection rates of Primordial Black Hole hyperbolic encounters. *Phys. Dark Univ.*, 21:61–69, 2018. doi:10.1016/j.dark.2018.06.001. (cit. on p. 22).

[34] Bernard Carr, Florian Kuhnel, and Marit Sandstad. Primordial Black Holes as Dark Matter. *Phys. Rev. D*, 94(8):083504, 2016. doi:10.1103/PhysRevD.94.083504. (cit. on p. 22).

[35] Albert Escrivà, Florian Kuhnel, and Yuichiro Tada. Primordial Black Holes. 11 2022. doi:10.48550/arXiv.2211.05767. (cit. on p. 22).

[36] Yacine Ali-Haïmoud, Ely D. Kovetz, and Marc Kamionkowski. Merger rate of primordial black-hole binaries. *Phys. Rev. D*, 96(12):123523, 2017. doi:10.1103/PhysRevD.96.123523. (cit. on p. 22).

[37] Steven L. Liebling and Carlos Palenzuela. Dynamical boson stars. *Living Rev. Rel.*, 26(1):1, 2023. doi:10.1007/s41114-023-00043-4. (cit. on p. 22).

[38] Andrew R. Liddle and Mark S. Madsen. The Structure and formation of boson stars. *Int. J. Mod. Phys. D*, 1:101–144, 1992. doi:10.1142/S0218271892000057. (cit. on p. 22).

[39] Jiajun Chen, Xiaolong Du, Erik W. Lentz, David J. E. Marsh, and Jens C. Niemeyer. New insights into the formation and growth of boson stars in dark matter halos. *Phys. Rev. D*, 104(8):083022, 2021. doi:10.1103/PhysRevD.104.083022. (cit. on p. 22).

[40] M. C. Begelman, R. D. Blandford, and M. J. Rees. Massive black hole binaries in active galactic nuclei. *Nature*, 287:307–309, 1980. doi:10.1038/287307a0. (cit. on p. 22).

[41] Antoine Klein et al. Science with the space-based interferometer eLISA: Supermassive black hole binaries. *Phys. Rev. D*, 93(2):024003, 2016. doi:10.1103/PhysRevD.93.024003. (cit. on p. 22).

[42] S. Sigurdsson and M. J. Rees. Capture of stellar mass compact objects by massive black holes in galactic cusps. *Mon. Not. Roy. Astron. Soc.*, 284:318, 1997. doi:10.1093/mnras/284.2.318. (cit. on p. 22).

[43] Pau Amaro-Seoane, Jonathan R. Gair, Marc Freitag, M. Coleman Miller, Ilya Mandel, Curt J. Cutler, and Stanislav Babak. Astrophysics, detection and science applications of intermediate- and extreme mass-ratio inspirals. *Class. Quant. Grav.*, 24: R113–R169, 2007. doi:10.1088/0264-9381/24/17/R01. (cit. on p. 22).

[44] Kostas D. Kokkotas and Bernd G. Schmidt. Quasinormal modes of stars and black holes. *Living Rev. Rel.*, 2:2, 1999. doi:10.12942/lrr-1999-2. (cit. on pp. 23 and 48).

[45] D. N. Spergel et al. First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters. *Astrophys. J. Suppl.*, 148: 175–194, 2003. doi:10.1086/377226. (cit. on p. 23).

[46] Rory Smith and Eric Thrane. Optimal Search for an Astrophysical Gravitational-Wave Background. *Phys. Rev. X*, 8(2):021019, 2018. doi:10.1103/PhysRevX.8.021019. (cit. on pp. 23 and 72).

[47] Steve Drasco and Eanna E. Flanagan. Detection methods for nonGaussian gravitational wave stochastic backgrounds. *Phys. Rev. D*, 67:082003, 2003. doi:10.1103/PhysRevD.67.082003. (cit. on p. 23).

[48] Brian D. Metzger. Kilonovae. *Living Rev. Rel.*, 23(1):1, 2020. doi:10.1007/s41114-019-0024-0. (cit. on pp. 23 and 117).

[49] B. P. Abbott et al. Multi-messenger Observations of a Binary Neutron Star Merger. *Astrophys. J. Lett.*, 848(2):L12, 2017. doi:10.3847/2041-8213/aa91c9. (cit. on pp. 23, 39, and 117).

[50] M. J. Graham et al. Candidate Electromagnetic Counterpart to the Binary Black Hole Merger Gravitational Wave Event S190521g. *Phys. Rev. Lett.*, 124(25):251102, 2020. doi:10.1103/PhysRevLett.124.251102. (cit. on p. 23).

[51] Cong Zhou, Lei Huang, Kangrou Guo, Ya-Ping Li, and Zhen Pan. Probing orbits of stellar mass objects deep in galactic nuclei with quasiperiodic eruptions. *Phys. Rev. D*, 109(10):103031, 2024. doi:10.1103/PhysRevD.109.103031. (cit. on p. 23).

[52] Vasileios Paschalidis and Nikolaos Stergioulas. Rotating Stars in Relativity. *Living Rev. Rel.*, 20(1):7, 2017. doi:10.1007/s41114-017-0008-x. (cit. on p. 24).

[53] Brynmor Haskell, D. I. Jones, and N. Andersson. Mountains on Neutron Stars: Accreted vs. Non-Accreted crusts. *Mon. Not. Roy. Astron. Soc.*, 373:1423–1439, 2006. doi:10.1111/j.1365-2966.2006.10998.x. (cit. on p. 24).

[54] M. Vigelius and A. Melatos. Resistive relaxation of a magnetically confined mountain on an accreting neutron star. *Mon. Not. Roy. Astron. Soc.*, 395, 1985. doi:10.1111/j.1365-2966.2009.14698.x. (cit. on p. 24).

[55] Fabian Gittins. Gravitational waves from neutron-star mountains. *Class. Quant. Grav.*, 41(4):043001, 2024. doi:10.1088/1361-6382/ad1c35. (cit. on p. 24).

[56] B. Abbott et al. Upper limits on gravitational wave emission from 78 radio pulsars. *Phys. Rev. D*, 76:042001, 2007. doi:10.1103/PhysRevD.76.042001. (cit. on p. 24).

[57] I. H. Stairs, A. G. Lyne, and S. L. Shemar. Evidence for free precession in a pulsar. *Nature*, 406:484–486, 2000. doi:10.1038/35020010. (cit. on p. 24).

[58] Benjamin J. Owen, Lee Lindblom, Curt Cutler, Bernard F. Schutz, Alberto Vecchio, and Nils Andersson. Gravitational waves from hot young rapidly rotating neutron stars. *Phys. Rev. D*, 58:084020, 1998. doi:10.1103/PhysRevD.58.084020. (cit. on p. 24).

[59] Kimberly C. B. New. Gravitational waves from gravitational collapse. *Living Rev. Rel.*, 6:2, 2003. doi:10.12942/lrr-2003-2. (cit. on p. 25).

[60] B. S. Sathyaprakash and B. F. Schutz. Physics, Astrophysics and Cosmology with Gravitational Waves. *Living Rev. Rel.*, 12:2, 2009. doi:10.12942/lrr-2009-2. (cit. on pp. 25 and 26).

[61] Masaru Shibata and Stuart L. Shapiro. Collapse of a rotating supermassive star to a supermassive black hole: Fully relativistic simulations. *Astrophys. J. Lett.*, 572: L39–L44, 2002. doi:10.1086/341516. (cit. on p. 25).

[62] Masaru Shibata, Thomas W. Baumgarte, and Stuart L. Shapiro. Stability and collapse of rapidly rotating, supramassive neutron stars: 3-D simulations in general relativity. *Phys. Rev. D*, 61:044012, 2000. doi:10.1103/PhysRevD.61.044012. (cit. on p. 25).

[63] Markus Rampp and H. Thomas Janka. Radiation hydrodynamics with neutrinos: Variable Eddington factor method for core collapse supernova simulations. *Astron. Astrophys.*, 396:361, 2002. doi:10.1051/0004-6361:20021398. (cit. on p. 25).

[64] Jens Hjorth et al. A Very energetic supernova associated with the gamma-ray burst of 29 March 2003. *Nature*, 423:847–850, 2003. doi:10.1038/nature01750. (cit. on p. 25).

[65] ChristianD. Ott. The Gravitational Wave Signature of Core-Collapse Supernovae. *Class. Quant. Grav.*, 26:063001, 2009. doi:10.1088/0264-9381/26/6/063001. (cit. on p. 25).

[66] Harald Dimmelmeier, Jose A. Font, and Ewald Muller. Relativistic simulations of rotational core collapse. 2. Collapse dynamics and gravitational radiation. *Astron. Astrophys.*, 393:523–542, 2002. doi:10.1051/0004-6361:20021053. (cit. on p. 25).

[67] Harald Dimmelmeier, Christian D. Ott, Hans-Thomas Janka, Andreas Marek, and Ewald Mueller. Generic Gravitational Wave Signals from the Collapse of Rotating Stellar Cores. *Phys. Rev. Lett.*, 98:251101, 2007. doi:10.1103/PhysRevLett.98.251101. (cit. on p. 25).

[68] Alan H. Guth. The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems. *Phys. Rev. D*, 23:347–356, 1981. doi:10.1103/PhysRevD.23.347. (cit. on p. 25).

[69] Rouzbeh Allahverdi, Robert Brandenberger, Francis-Yan Cyr-Racine, and Anupam Mazumdar. Reheating in Inflationary Cosmology: Theory and Applications. *Ann. Rev. Nucl. Part. Sci.*, 60:27–51, 2010. doi:10.1146/annurev.nucl.012809.104511. (cit. on p. 25).

[70] Michael S. Turner and Frank Wilczek. Relic gravitational waves and extended inflation. *Phys. Rev. Lett.*, 65:3080–3083, 1990. doi:10.1103/PhysRevLett.65.3080. (cit. on p. 25).

[71] A. Vilenkin. Gravitational radiation from cosmic strings. *Phys. Lett. B*, 107:47–50, 1981. doi:10.1016/0370-2693(81)91144-8. (cit. on p. 25).

[72] B. P. Abbott et al. Search for the isotropic stochastic background using data from Advanced LIGO's second observing run. *Phys. Rev. D*, 100(6):061101, 2019. doi:10.1103/PhysRevD.100.061101. (cit. on p. 26).

[73] J. Weber. Detection and generation of gravitational waves. *Phys. Rev.*, 117:306–313, Jan 1960. doi:10.1103/PhysRev.117.306. URL https://link.aps.org/doi/10.1103/PhysRev.117.306. (cit. on p. 26).

[74] J. Weber. Evidence for discovery of gravitational radiation. *Phys. Rev. Lett.*, 22:1320–1324, Jun 1969. doi:10.1103/PhysRevLett.22.1320. URL https://link.aps.org/doi/10.1103/PhysRevLett.22.1320. (cit. on p. 26).

[75] R. W. P. Drever, J. Hough, R. Bland, and G. W. Lessnoff. Search for Short Bursts of Gravitational Radiation. *Nature*, 246(5432):340–344, December 1973. doi:10.1038/246340a0. (cit. on p. 26).

[76] James L. Levine and Richard L. Garwin. Absence of gravity-wave signals in a bar at 1695 hz. *Phys. Rev. Lett.*, 31:173–176, Jul 1973. doi:10.1103/PhysRevLett.31.173. URL https://link.aps.org/doi/10.1103/PhysRevLett.31.173. (cit. on p. 26).

[77] H. Billing, P. Kafka, K. Maischberger, F. Meyer, and W. Winkler. Results of the Munich-Frascati gravitational-wave experiment. *Nuovo Cimento Lettere*, 12:111–116, January 1975. doi:10.1007/BF02790471. (cit. on p. 26).

[78] Hiromasa Hirakawa and Kazumichi Narihara. Search for gravitational radiation at 145 hz. *Phys. Rev. Lett.*, 35:330–334, Aug 1975. doi:10.1103/PhysRevLett.35.330. URL https://link.aps.org/doi/10.1103/PhysRevLett.35.330. (cit. on p. 26).

[79] V.B. Braginsky, A.B. Manukin, E.I. Popov, and V.N. Rudenko. The search for gravitational radiation of non-terrestrial origin. *Physics Letters A*, 45(4):271–272, 1973. ISSN 0375-9601. doi:https://doi.org/10.1016/0375-9601(73)90074-1. URL https://www.sciencedirect.com/science/article/pii/0375960173900741. (cit. on p. 26).

[80] M. E. Gertsenshteĭn and V. I. Pustovoĭt. On the Detection of Low-Frequency Gravitational Waves. *Soviet Journal of Experimental and Theoretical Physics*, 16:433, January 1963. (cit. on p. 26).

[81] H. Luck. The GEO-600 project. *Class. Quant. Grav.*, 14:1471–1476, 1997. doi:10.1088/0264-9381/14/6/012. (cit. on p. 26).

[82] B. Abbott et al. Joint ligo and tama300 search for gravitational waves from inspiralling neutron star binaries. *Phys. Rev. D*, 73:102002, May 2006. doi:10.1103/PhysRevD.73.102002. URL https://link.aps.org/doi/10.1103/PhysRevD.73.102002. (cit. on p. 26).

[83] J. Aasi et al. Improved Upper Limits on the Stochastic Gravitational-Wave Background from 2009–2010 LIGO and Virgo Data. *Phys. Rev. Lett.*, 113(23):231101, 2014. doi:10.1103/PhysRevLett.113.231101. (cit. on p. 26).

[84] Gabriella Agazie et al. The NANOGrav 15 yr Data Set: Evidence for a Gravitational-wave Background. *Astrophys. J. Lett.*, 951(1):L8, 2023. doi:10.3847/2041-8213/acdac6. (cit. on pp. 27 and 42).

[85] J. Antoniadis et al. The second data release from the European Pulsar Timing Array - III. Search for gravitational wave signals. *Astron. Astrophys.*, 678:A50, 2023. doi:10.1051/0004-6361/202346844. (cit. on pp. 27 and 42).

[86] Daniel J. Reardon et al. Search for an Isotropic Gravitational-wave Background with the Parkes Pulsar Timing Array. *Astrophys. J. Lett.*, 951(1):L6, 2023. doi:10.3847/2041-8213/acdd02. (cit. on pp. 27 and 42).

[87] Heng Xu et al. Searching for the Nano-Hertz Stochastic Gravitational Wave Background with the Chinese Pulsar Timing Array Data Release I. *Res. Astron. Astrophys.*, 23(7):075024, 2023. doi:10.1088/1674-4527/acdfa5. (cit. on pp. 27 and 42).

[88] J. Aasi et al. Advanced LIGO. *Class. Quant. Grav.*, 32:074001, 2015. doi:10.1088/0264-9381/32/7/074001. (cit. on pp. 27, 28, 33, and 35).

[89] G. Cagnoli et al. Silica suspension and coating developments for Advanced LIGO. *J. Phys. Conf. Ser.*, 32:386–392, 2006. doi:10.1088/1742-6596/32/1/059. (cit. on p. 27).

[90] L. Carbone et al. Sensors and Actuators for the Advanced LIGO Mirror Suspensions. *Class. Quant. Grav.*, 29:115005, 2012. doi:10.1088/0264-9381/29/11/115005. (cit. on p. 27).

[91] P. Kwee, C. Bogan, K. Danzmann, M. Frede, H. Kim, P. King, J. Pöld, O. Puncken, R. L. Savage, F. Seifert, P. Wessels, L. Winkelmann, and B. Willke. Stabilized high-power laser system for the gravitational wave detector advanced ligo. *Opt. Express*, 20(10):10617–10634, May 2012. doi:10.1364/OE.20.010617. URL https://opg.optica.org/oe/abstract.cfm?URI=oe-20-10-10617. (cit. on p. 27).

[92] B. P. Abbott et al. LIGO: The Laser interferometer gravitational-wave observatory. *Rept. Prog. Phys.*, 72:076901, 2009. doi:10.1088/0034-4885/72/7/076901. (cit. on pp. 27 and 28).

[93] L. Pinard, C. Michel, B. Sassolas, L. Balzarini, J. Degallaix, V. Dolique, R. Flaminio, D. Forest, M. Granata, B. Lagrange, N. Straniero, J. Teillon, and G. Cagnoli. Mirrors used in the ligo interferometers for first detection of gravitational waves. *Appl. Opt.*, 56(4):C11–C15, Feb 2017. doi:10.1364/AO.56.000C11. URL https://opg.optica.org/ao/abstract.cfm?URI=ao-56-4-C11. (cit. on p. 27).

[94] Massimo Granata et al. Amorphous optical coatings of present gravitational-wave interferometers. *Class. Quant. Grav.*, 37(9):095004, 2020. doi:10.1088/1361-6382/ab77e9. (cit. on p. 27).

[95] W. A. Edelstein, J. Hough, J. R. Pugh, and W. Martin. LETTER TO THE EDITOR: Limits to the measurement of displacement in an interferometric gravitational radiation detector. *Journal of Physics E Scientific Instruments*, 11(7):710–712, July 1978. doi:10.1088/0022-3735/11/7/030. (cit. on p. 28).

[96] Muzammil A. Arain and Guido Mueller. Design of the advanced ligo recycling cavities. *Opt. Express*, 16(14):10018–10032, Jul 2008. doi:10.1364/OE.16.010018. URL https://opg.optica.org/oe/abstract.cfm?URI=oe-16-14-10018. (cit. on p. 28).

[97] Brian J. Meers. Recycling in laser-interferometric gravitational-wave detectors. *Phys. Rev. D*, 38:2317–2326, Oct 1988. doi:10.1103/PhysRevD.38.2317. URL https://link.aps.org/doi/10.1103/PhysRevD.38.2317. (cit. on p. 28).

[98] Matthew Pitkin, Stuart Reid, Sheila Rowan, and Jim Hough. Gravitational Wave Detection by Interferometry (Ground and Space). *Living Rev. Rel.*, 14:5, 2011. doi:10.12942/lrr-2011-5. (cit. on p. 28).

[99] J. Aasi et al. Enhancing the sensitivity of the LIGO gravitational wave detector by using squeezed states of light. *Nature Photon.*, 7:613–619, 2013. doi:10.1038/nphoton.2013.177. (cit. on p. 28).

[100] Haocun Yu et al. Quantum correlations between light and the kilogram-mass mirrors of LIGO. *Nature*, 583(7814):43–47, 2020. doi:10.1038/s41586-020-2420-8. (cit. on pp. 28 and 34).

[101] M. Tse et al. Quantum-Enhanced Advanced LIGO Detectors in the Era of Gravitational-Wave Astronomy. *Phys. Rev. Lett.*, 123(23):231107, 2019. doi:10.1103/PhysRevLett.123.231107. (cit. on p. 28).

[102] F. Acernese et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. *Class. Quant. Grav.*, 32(2):024001, 2015. doi:10.1088/0264-9381/32/2/024001. (cit. on pp. 28 and 39).

[103] T. Akutsu et al. KAGRA: 2.5 Generation Interferometric Gravitational Wave Detector. *Nature Astron.*, 3(1):35–40, 2019. doi:10.1038/s41550-018-0658-y. (cit. on p. 28).

[104] John T. Whelan. The geometry of gravitational wave detection, 2013. URL https://dcc.ligo.org/LIGO-T1300666/public. (cit. on p. 28).

[105] B. P. Abbott et al. Calibration of the Advanced LIGO detectors for the discovery of the binary black-hole merger GW150914. *Phys. Rev. D*, 95(6):062003, 2017. doi:10.1103/PhysRevD.95.062003. (cit. on p. 29).

[106] Piotr Jaranowski, Andrzej Krolak, and Bernard F. Schutz. Data analysis of gravitational - wave signals from spinning neutron stars. 1. The Signal and its detection. *Phys. Rev. D*, 58:063001, 1998. doi:10.1103/PhysRevD.58.063001. (cit. on pp. 30, 32, 118, and 125).

[107] Warren G. Anderson, Patrick R. Brady, Jolien D. E. Creighton, and Eanna E. Flanagan. An Excess power statistic for detection of burst sources of gravitational radiation. *Phys. Rev. D*, 63:042003, 2001. doi:10.1103/PhysRevD.63.042003. (cit. on pp. 30 and 32).

[108] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. LVK Algorithm Library - LALSuite. Free software (GPL), 2018. (cit. on pp. 30 and 33).

[109] Gregory Ashton et al. BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy. *Astrophys. J. Suppl.*, 241(2):27, 2019. doi:10.3847/1538-4365/ab06fc. (cit. on pp. 30 and 106).

[110] Md Arif Shaikh, Vijay Varma, Harald P. Pfeiffer, Antoni Ramos-Buades, and Maarten van de Meent. Defining eccentricity for gravitational wave astronomy. *Phys. Rev. D*, 108(10):104007, 2023. doi:10.1103/PhysRevD.108.104007. (cit. on pp. 30 and 43).

[111] Lee S. Finn. Detection, measurement and gravitational radiation. *Phys. Rev. D*, 46: 5236–5249, 1992. doi:10.1103/PhysRevD.46.5236. (cit. on pp. 33, 36, 63, and 67).

[112] Linqing Wen and Yanbei Chen. Geometrical Expression for the Angular Resolution of a Network of Gravitational-Wave Detectors. *Phys. Rev. D*, 81:082001, 2010. doi:10.1103/PhysRevD.81.082001. (cit. on pp. 33, 118, and 128).

[113] Beverly K. Berger and LIGO Scientific Collaboration. Tracking down the origins of Advanced LIGO noise: some examples. In *APS April Meeting Abstracts*, volume 2018 of *APS Meeting Abstracts*, page S14.004, January 2018. (cit. on p. 35).

[114] C. J. Moore, R. H. Cole, and C. P. L. Berry. Gravitational-wave sensitivity curves. *Class. Quant. Grav.*, 32(1):015014, 2015. doi:10.1088/0264-9381/32/1/015014. (cit. on p. 35).

[115] Benjamin P Abbott et al. A guide to LIGO–Virgo detector noise and extraction of transient gravitational-wave signals. *Class. Quant. Grav.*, 37(5):055002, 2020. doi:10.1088/1361-6382/ab685e. (cit. on p. 36).

[116] B. P. Abbott et al. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Phys. Rev. Lett.*, 119(16):161101, 2017. doi:10.1103/PhysRevLett.119.161101. (cit. on pp. 36, 39, and 117).

[117] Ethan Payne, Sophie Hourihane, Jacob Golomb, Rhiannon Udall, Richard Udall, Derek Davis, and Katerina Chatziioannou. Curious case of GW200129: Interplay between spin-precession inference and data-quality issues. *Phys. Rev. D*, 106(10): 104017, 2022. doi:10.1103/PhysRevD.106.104017. (cit. on p. 36).

[118] Michael Zevin et al. Gravity Spy: Integrating Advanced LIGO Detector Characterization, Machine Learning, and Citizen Science. *Class. Quant. Grav.*, 34(6):064003, 2017. doi:10.1088/1361-6382/aa5cea. (cit. on pp. 36 and 72).

[119] D. M. MacLeod, S. Fairhurst, B. Hughey, A. P. Lundgren, L. Pekowsky, J. Rollins, and J. R. Smith. Reducing the effect of seismic noise in LIGO searches by targeted veto generation. *Class. Quant. Grav.*, 29:055006, 2012. doi:10.1088/0264-9381/29/5/055006. (cit. on p. 36).

[120] S. Soni et al. LIGO Detector Characterization in the first half of the fourth Observing run, 9 2024. (cit. on p. 36).

[121] Miriam Cabero et al. Blip glitches in Advanced LIGO data. *Class. Quant. Grav.*, 36(15):15, 2019. doi:10.1088/1361-6382/ab2e14. (cit. on p. 36).

[122] Daniel George, Hongyu Shen, and E. A. Huerta. Classification and unsupervised clustering of LIGO data with Deep Transfer Learning. *Phys. Rev. D*, 97(10):101501, 2018. doi:10.1103/PhysRevD.97.101501. (cit. on p. 36).

[123] S. Soni et al. Discovering features in gravitational-wave data through detector characterization, citizen science and machine learning. *Class. Quant. Grav.*, 38(19): 195016, 2021. doi:10.1088/1361-6382/ac1ccb. (cit. on p. 36).

[124] C. Kittel. *Elementary Statistical Physics*. Dover Books on Physics Series. Dover Publications, 2004. ISBN 9780486435145. URL https://books.google.co.uk/books?id=5sd9SAoRjgQC. (cit. on p. 36).

[125] B. P. Abbott et al. GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs. *Phys. Rev. X*, 9(3):031040, 2019. doi:10.1103/PhysRevX.9.031040. (cit. on pp. 39 and 101).

[126] R. Abbott et al. GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run. *Phys. Rev. X*, 11:021053, 2021. doi:10.1103/PhysRevX.11.021053. (cit. on pp. 39, 49, and 102).

[127] R. Abbott et al. GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run. *Phys. Rev. D*, 109(2):022001, 2024. doi:10.1103/PhysRevD.109.022001. (cit. on pp. 39, 49, 78, 101, 102, 103, and 108).

[128] R. Abbott et al. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo during the Second Part of the Third Observing Run. *Phys. Rev. X*, 13(4): 041039, 2023. doi:10.1103/PhysRevX.13.041039. (cit. on pp. 39, 78, 97, 101, 102, 103, and 108).

[129] A. G. Abac et al. Observation of Gravitational Waves from the Coalescence of a $2.5 - 4.5\ M_\odot$ Compact Object and a Neutron Star. 4 2024. (cit. on p. 39).

[130] Ryutaro Takahashi and the TAMA Collaboration. Status of tama300. *Classical and Quantum Gravity*, 21(5):S403, feb 2004. doi:10.1088/0264-9381/21/5/004. URL https://dx.doi.org/10.1088/0264-9381/21/5/004. (cit. on p. 39).

[131] K. L. Dooley et al. GEO 600 and the GEO-HF upgrade program: successes and challenges. *Class. Quant. Grav.*, 33:075009, 2016. doi:10.1088/0264-9381/33/7/075009. (cit. on p. 39).

[132] R. Abbott et al. First joint observation by the underground gravitational-wave detector KAGRA with GEO 600. *PTEP*, 2022(6):063F01, 2022. doi:10.1093/ptep/ptac073. (cit. on p. 39).

[133] B. 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. doi:10.3847/2041-8213/aa920c. (cit. on p. 39).

[134] A. Albert et al. Search for High-energy Neutrinos from Binary Neutron Star Merger GW170817 with ANTARES, IceCube, and the Pierre Auger Observatory. *Astrophys. J. Lett.*, 850(2):L35, 2017. doi:10.3847/2041-8213/aa9aed. (cit. on p. 39).

[135] B. P. Abbott et al. GW170817: Measurements of neutron star radii and equation of state. *Phys. Rev. Lett.*, 121(16):161101, 2018. doi:10.1103/PhysRevLett.121.161101. (cit. on pp. 39 and 117).

[136] B. P. Abbott et al. Properties of the binary neutron star merger GW170817. *Phys. Rev. X*, 9(1):011001, 2019. doi:10.1103/PhysRevX.9.011001. (cit. on pp. 39, 43, 106, 117, 129, and 163).

[137] R. Abbott et al. GW190521: A Binary Black Hole Merger with a Total Mass of $150 M_\odot$. *Phys. Rev. Lett.*, 125(10):101102, 2020. doi:10.1103/PhysRevLett.125.101102. (cit. on p. 39).

[138] R. Abbott et al. GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object. *Astrophys. J. Lett.*, 896(2):L44, 2020. doi:10.3847/2041-8213/ab960f. (cit. on p. 39).

[139] Charlie Hoy, Cameron Mills, and Stephen Fairhurst. Evidence for subdominant multipole moments and precession in merging black-hole-binaries from GWTC-2.1. *Phys. Rev. D*, 106(2):023019, 2022. doi:10.1103/PhysRevD.106.023019. (cit. on p. 39).

[140] Mark Hannam et al. General-relativistic precession in a black-hole binary. *Nature*, 610(7933):652–655, 2022. doi:10.1038/s41586-022-05212-z. (cit. on pp. 39 and 44).

[141] Isobel M. Romero-Shaw, Paul D. Lasky, Eric Thrane, and Juan Calderon Bustillo. GW190521: orbital eccentricity and signatures of dynamical formation in a binary black hole merger signal. *Astrophys. J. Lett.*, 903(1):L5, 2020. doi:10.3847/2041-8213/abbe26. (cit. on p. 39).

[142] Nihar Gupte et al. Evidence for eccentricity in the population of binary black holes observed by LIGO-Virgo-KAGRA. 4 2024. doi:10.48550/arXiv.2404.14286. (cit. on pp. 39, 73, and 141).

[143] Roberto Cotesta, Gregorio Carullo, Emanuele Berti, and Vitor Cardoso. Analysis of Ringdown Overtones in GW150914. *Phys. Rev. Lett.*, 129(11):111102, 2022. doi:10.1103/PhysRevLett.129.111102. (cit. on p. 39).

[144] Maximiliano Isi and Will M. Farr. Comment on "Analysis of Ringdown Overtones in GW150914". *Phys. Rev. Lett.*, 131(16):169001, 2023. doi:10.1103/PhysRevLett.131.169001. (cit. on p. 39).

[145] Gregorio Carullo, Roberto Cotesta, Emanuele Berti, and Vitor Cardoso. Reply to Comment on "Analysis of Ringdown Overtones in GW150914". *Phys. Rev. Lett.*, 131:169002, 2023. doi:10.1103/PhysRevLett.131.169002. (cit. on p. 39).

[146] Yi-Fan Wang, Collin D. Capano, Jahed Abedi, Shilpa Kastha, Badri Krishnan, Alex B. Nielsen, Alexander H. Nitz, and Julian Westerweck. A frequency-domain perspective on GW150914 ringdown overtone. 10 2023. doi:10.48550/arXiv.2310.19645. (cit. on p. 39).

[147] Vijay Varma, Sylvia Biscoveanu, Tousif Islam, Feroz H. Shaik, Carl-Johan Haster, Maximiliano Isi, Will M. Farr, Scott E. Field, and Salvatore Vitale. Evidence of Large Recoil Velocity from a Black Hole Merger Signal. *Phys. Rev. Lett.*, 128(19): 191102, 2022. doi:10.1103/PhysRevLett.128.191102. (cit. on p. 39).

[148] B. P. Abbott et al. Binary Black Hole Population Properties Inferred from the First and Second Observing Runs of Advanced LIGO and Advanced Virgo. *Astrophys. J. Lett.*, 882(2):L24, 2019. doi:10.3847/2041-8213/ab3800. (cit. on p. 40).

[149] R. Abbott et al. Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog. *Astrophys. J. Lett.*, 913(1):L7, 2021. doi:10.3847/2041-8213/abe949. (cit. on pp. 40 and 85).

[150] R. Abbott et al. Population of Merging Compact Binaries Inferred Using Gravitational Waves through GWTC-3. *Phys. Rev. X*, 13(1):011048, 2023. doi:10.1103/PhysRevX.13.011048. (cit. on pp. 40 and 49).

[151] B. P. Abbott et al. A gravitational-wave standard siren measurement of the Hubble constant. *Nature*, 551(7678):85–88, 2017. doi:10.1038/nature24471. (cit. on p. 40).

[152] R. Abbott et al. Constraints on the Cosmic Expansion History from GWTC–3. *Astrophys. J.*, 949(2):76, 2023. doi:10.3847/1538-4357/ac74bb. (cit. on pp. 40 and 91).

[153] M. Soares-Santos 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. *Astrophys. J. Lett.*, 876(1):L7, 2019. doi:10.3847/2041-8213/ab14f1. (cit. on p. 40).

[154] B. P. Abbott et al. A Gravitational-wave Measurement of the Hubble Constant Following the Second Observing Run of Advanced LIGO and Virgo. *Astrophys. J.*, 909(2):218, 2021. doi:10.3847/1538-4357/abdcb7. (cit. on p. 40).

[155] B. P. Abbott et al. Tests of general relativity with GW150914. *Phys. Rev. Lett.*, 116 (22):221101, 2016. doi:10.1103/PhysRevLett.116.221101. [Erratum: Phys.Rev.Lett. 121, 129902 (2018)]. (cit. on pp. 40 and 78).

[156] B. P. Abbott et al. Tests of General Relativity with GW170817. *Phys. Rev. Lett.*, 123(1):011102, 2019. doi:10.1103/PhysRevLett.123.011102. (cit. on pp. 40 and 78).

[157] B. P. Abbott et al. Tests of General Relativity with the Binary Black Hole Signals from the LIGO-Virgo Catalog GWTC-1. *Phys. Rev. D*, 100(10):104036, 2019. doi:10.1103/PhysRevD.100.104036. (cit. on pp. 40 and 78).

[158] R. Abbott et al. Tests of general relativity with binary black holes from the second LIGO-Virgo gravitational-wave transient catalog. *Phys. Rev. D*, 103(12):122002, 2021. doi:10.1103/PhysRevD.103.122002. (cit. on pp. 40 and 78).

[159] R. Abbott et al. Constraints on dark photon dark matter using data from LIGO's and Virgo's third observing run. *Phys. Rev. D*, 105(6):063030, 2022.

doi:10.1103/PhysRevD.105.063030. [Erratum: Phys.Rev.D 109, 089902 (2024)]. (cit. on p. 40).

[160] A. G. Abac et al. Ultralight vector dark matter search using data from the KAGRA O3GK run. 3 2024. doi:10.48550/arXiv.2403.03004. (cit. on p. 40).

[161] R. Abbott et al. All-sky search for short gravitational-wave bursts in the third Advanced LIGO and Advanced Virgo run. *Phys. Rev. D*, 104(12):122004, 2021. doi:10.1103/PhysRevD.104.122004. (cit. on p. 40).

[162] R. Abbott et al. All-sky search for continuous gravitational waves from isolated neutron stars in the early O3 LIGO data. *Phys. Rev. D*, 104(8):082004, 2021. doi:10.1103/PhysRevD.104.082004. (cit. on p. 40).

[163] R. Abbott et al. All-sky search for long-duration gravitational-wave bursts in the third Advanced LIGO and Advanced Virgo run. *Phys. Rev. D*, 104(10):102001, 2021. doi:10.1103/PhysRevD.104.102001. (cit. on p. 40).

[164] M. Saleem et al. The science case for LIGO-India. *Class. Quant. Grav.*, 39(2): 025004, 2022. doi:10.1088/1361-6382/ac3b99. (cit. on p. 40).

[165] B. Abbott et al. Setting upper limits on the strength of periodic gravitational waves using the first science data from the GEO 600 and LIGO detectors. *Phys. Rev. D*, 69:082004, 2004. doi:10.1103/PhysRevD.69.082004. (cit. on p. 40).

[166] M. Punturo et al. The Einstein Telescope: A third-generation gravitational wave observatory. *Class. Quant. Grav.*, 27:194002, 2010. doi:10.1088/0264-9381/27/19/194002. (cit. on pp. 40, 78, 86, and 117).

[167] Marica Branchesi et al. Science with the Einstein Telescope: a comparison of different designs. *JCAP*, 07:068, 2023. doi:10.1088/1475-7516/2023/07/068. (cit. on pp. 40, 117, and 133).

[168] David Reitze and et al. Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO. (arXiv:1907.04833), July 2019. doi:10.48550/arXiv.1907.04833. URL `http://arxiv.org/abs/1907.04833`. (cit. on pp. 40, 78, and 117).

[169] Matthew Evans et al. Cosmic Explorer: A Submission to the NSF MPSAC ngGW Subcommittee. 6 2023. doi:10.48550/arXiv.2306.13745. (cit. on pp. 40 and 117).

[170] Michele Maggiore et al. Science Case for the Einstein Telescope. *JCAP*, 03:050, 2020. doi:10.1088/1475-7516/2020/03/050. (cit. on p. 40).

[171] Ssohrab Borhanian and B. S. Sathyaprakash. Listening to the Universe with Next Generation Ground-Based Gravitational-Wave Detectors. 2 2022. doi:10.48550/arXiv.2202.11048. (cit. on pp. 40, 75, 81, 117, 133, and 161).

[172] Teng Zhang, Huan Yang, Denis Martynov, Patricia Schmidt, and Haixing Miao. Gravitational-Wave Detector for Postmerger Neutron Stars: Beyond the Quantum Loss Limit of the Fabry-Perot-Michelson Interferometer. *Phys. Rev. X*, 13(2):021019, 2023. doi:10.1103/PhysRevX.13.021019. (cit. on p. 40).

[173] B. P. Abbott et al. Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Rev. Rel.*, 21(1): 3, 2018. doi:10.1007/s41114-020-00026-9. (cit. on p. 41).

[174] Benjamin P Abbott et al. Exploring the Sensitivity of Next Generation Gravitational Wave Detectors. *Class. Quant. Grav.*, 34(4):044001, 2017. doi:10.1088/1361-6382/aa51f4. (cit. on pp. 41 and 86).

[175] Pau Amaro-Seoane et al. Laser interferometer space antenna. 2017. doi:10.48550/arXiv.1702.00786. (cit. on p. 40).

[176] Lisa pathfinder. 2019. doi:10.48550/arXiv.1903.08924. (cit. on p. 41).

[177] Jun Luo et al. TianQin: a space-borne gravitational wave detector. *Class. Quant. Grav.*, 33(3):035010, 2016. doi:10.1088/0264-9381/33/3/035010. (cit. on p. 41).

[178] Wen-Rui Hu and Yue-Liang Wu. The Taiji Program in Space for gravitational wave physics and the nature of gravity. *Natl. Sci. Rev.*, 4(5):685–686, 2017. doi:10.1093/nsr/nwx116. (cit. on p. 41).

[179] Jun Luo et al. The first round result from the TianQin-1 satellite. *Class. Quant. Grav.*, 37(18):185013, 2020. doi:10.1088/1361-6382/aba66a. (cit. on p. 41).

[180] Yue-Liang Wu and et al. China's first step towards probing the expanding universe and the nature of gravity using a space borne gravitational wave antenna. *Communications Physics*, 4(1):34, February 2021. ISSN 2399-3650. doi:10.1038/s42005-021-00529-z. URL `https://doi.org/10.1038/s42005-021-00529-z`. (cit. on p. 41).

[181] S. Kawamura et al. The Japanese space gravitational wave antenna DECIGO. *Class. Quant. Grav.*, 23:S125–S132, 2006. doi:10.1088/0264-9381/23/8/S17. (cit. on p. 41).

[182] M. V. Sazhin. Opportunities for detecting ultralong gravitational waves. *Soviet Astronomy*, 22:36–38, February 1978. (cit. on p. 42).

[183] S. Detweiler. Pulsar timing measurements and the search for gravitational waves. *Astrophys. J.*, 234:1100–1104, December 1979. doi:10.1086/157593. (cit. on p. 42).

[184] R. W. Hellings and G. S. Downs. Upper limits on the isotropic gravitational radiation background from pulsar timing analysis. *Astrophys. J. Lett.*, 265:L39–L42, February 1983. doi:10.1086/183954. (cit. on p. 42).

[185] Remi Geiger. Future Gravitational Wave Detectors Based on Atom Interferometry. pages 285–313, 2017. doi:10.1142/9789813141766_0008. (cit. on p. 42).

[186] Dongfeng Gao, Jin Wang, and Mingsheng Zhan. Atomic Interferometric Gravitational-wave Space Observatory (AIGSO). *Commun. Theor. Phys.*, 69(1): 37, 2018. doi:10.1088/0253-6102/69/1/37. (cit. on p. 42).

[187] Atsushi Nishizawa et al. Neutron displacement noise-free interferometer for gravitational-wave detection. *Phys. Rev. D*, 105(12):124017, 2022. doi:10.1103/PhysRevD.105.124017. (cit. on p. 42).

[188] Parameswaran Ajith et al. The Lunar Gravitational-wave Antenna: Mission Studies and Science Case. 4 2024. doi:10.48550/arXiv.2404.09181. (cit. on p. 42).

[189] Michael Coughlin and Jan Harms. Constraining the gravitational wave energy density of the Universe using Earth's ring. *Phys. Rev. D*, 90(4):042005, 2014. doi:10.1103/PhysRevD.90.042005. (cit. on p. 42).

[190] Hong Li et al. Probing Primordial Gravitational Waves: Ali CMB Polarization Telescope. *Natl. Sci. Rev.*, 6(1):145–154, 2019. doi:10.1093/nsr/nwy019. (cit. on p. 42).

[191] Laura G. Book and Eanna E. Flanagan. Astrometric Effects of a Stochastic Gravitational Wave Background. *Phys. Rev. D*, 83:024024, 2011. doi:10.1103/PhysRevD.83.024024. (cit. on p. 42).

[192] J. W. Armstrong. Low-frequency gravitational wave searches using spacecraft doppler tracking. *Living Reviews in Relativity*, 9(1):1, January 2006. ISSN 1433-8351. doi:10.12942/lrr-2006-1. URL `https://doi.org/10.12942/lrr-2006-1`. (cit. on p. 42).

[193] Benjamin Farr, Evan Ochsner, Will M. Farr, and Richard O'Shaughnessy. A more effective coordinate system for parameter estimation of precessing compact binaries from gravitational waves. *Phys. Rev. D*, 90(2):024018, 2014. doi:10.1103/PhysRevD.90.024018. (cit. on p. 44).

[194] E. Poisson and C.M. Will. *Gravity: Newtonian, Post-Newtonian, Relativistic*. Cambridge University Press, 2014. ISBN 9781107032866. URL https://books.google.co.uk/books?id=PZ5cAwAAQBAJ. (cit. on p. 44).

[195] 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. doi:10.1103/PhysRevD.49.6274. URL https://link.aps.org/doi/10.1103/PhysRevD.49.6274. (cit. on p. 44).

[196] Thibault Damour. Coalescence of two spinning black holes: An effective one-body approach. *Phys. Rev. D*, 64:124013, Nov 2001. doi:10.1103/PhysRevD.64.124013. URL https://link.aps.org/doi/10.1103/PhysRevD.64.124013. (cit. on p. 44).

[197] 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. *Phys. Rev. Lett.*, 106:241101, Jun 2011. doi:10.1103/PhysRevLett.106.241101. URL https://link.aps.org/doi/10.1103/PhysRevLett.106.241101. (cit. on p. 44).

[198] L. Santamaría, F. Ohme, P. Ajith, B. Brügmann, N. Dorband, M. Hannam, S. Husa, P. Mösta, D. Pollney, C. Reisswig, E. L. Robinson, J. Seiler, and B. Krishnan. Matching post-newtonian and numerical relativity waveforms: Systematic errors and a new phenomenological model for nonprecessing black hole binaries. *Phys. Rev. D*, 82:064016, Sep 2010. doi:10.1103/PhysRevD.82.064016. URL https://link.aps.org/doi/10.1103/PhysRevD.82.064016. (cit. on p. 44).

[199] Curt Cutler and Eanna E. Flanagan. Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral wave form? *Phys. Rev. D*, 49:2658–2697, 1994. doi:10.1103/PhysRevD.49.2658. (cit. on pp. 44 and 47).

[200] Eric Poisson and Clifford M. Will. Gravitational waves from inspiraling compact binaries: Parameter estimation using second postNewtonian wave forms. *Phys. Rev. D*, 52:848–855, 1995. doi:10.1103/PhysRevD.52.848. (cit. on pp. 44 and 47).

[201] Mark Hannam, Patricia Schmidt, Alejandro Bohé, Leïla Haegel, Sascha Husa, Frank Ohme, Geraint Pratten, and Michael Pürrer. Simple Model of Complete Precessing Black-Hole-Binary Gravitational Waveforms. *Phys. Rev. Lett.*, 113(15):151101, 2014. doi:10.1103/PhysRevLett.113.151101. (cit. on p. 44).

[202] Patricia Schmidt, Frank Ohme, and Mark Hannam. Towards models of gravitational waveforms from generic binaries II: Modelling precession effects with

a single effective precession parameter. *Phys. Rev. D*, 91(2):024043, 2015. doi:10.1103/PhysRevD.91.024043. (cit. on p. 44).

[203] B. P. Abbott et al. Properties of the Binary Black Hole Merger GW150914. *Phys. Rev. Lett.*, 116(24):241102, 2016. doi:10.1103/PhysRevLett.116.241102. (cit. on p. 44).

[204] Salvatore Vitale, Ryan Lynch, Vivien Raymond, Riccardo Sturani, John Veitch, and Philp Graff. Parameter estimation for heavy binary-black holes with networks of second-generation gravitational-wave detectors. *Phys. Rev. D*, 95(6):064053, 2017. doi:10.1103/PhysRevD.95.064053. (cit. on p. 44).

[205] 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. doi:10.1103/PhysRevD.89.103012. URL https://link.aps.org/doi/10.1103/PhysRevD.89.103012. (cit. on p. 46).

[206] Luc Blanchet. Gravitational Radiation from Post-Newtonian Sources and Inspiralling Compact Binaries. *Living Rev. Rel.*, 17:2, 2014. doi:10.12942/lrr-2014-2. (cit. on pp. 46 and 47).

[207] Alessandra Buonanno, Bala Iyer, Evan Ochsner, Yi Pan, and B. S. Sathyaprakash. Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors. *Phys. Rev. D*, 80:084043, 2009. doi:10.1103/PhysRevD.80.084043. (cit. on pp. 46 and 131).

[208] Lawrence E. Kidder, Clifford M. Will, and Alan G. Wiseman. Spin effects in the inspiral of coalescing compact binaries. *Phys. Rev. D*, 47(10):R4183–R4187, 1993. doi:10.1103/PhysRevD.47.R4183. (cit. on p. 47).

[209] 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. doi:10.1103/PhysRevLett.74.3515. URL https://link.aps.org/doi/10.1103/PhysRevLett.74.3515. (cit. on p. 47).

[210] Philippe Grandclement and Jerome Novak. Spectral methods for numerical relativity. *Living Rev. Rel.*, 12:1, 2009. doi:10.12942/lrr-2009-1. (cit. on p. 47).

[211] Nigel T. Bishop and Luciano Rezzolla. Extraction of Gravitational Waves in Numerical Relativity. *Living Rev. Rel.*, 19:2, 2016. doi:10.1007/s41114-016-0001-9. (cit. on p. 47).

[212] S. A. Hughes. Gravitational waves from merging compact binaries. *Ann. Rev. Astron. Astrophys.*, 47:107–157, 2009. doi:10.1146/annurev-astro-082708-101711. (cit. on p. 47).

[213] Frans Pretorius. Numerical relativity using a generalized harmonic decomposition. *Class. Quant. Grav.*, 22:425–452, 2005. doi:10.1088/0264-9381/22/2/014. (cit. on p. 47).

[214] Frans Pretorius. Evolution of binary black hole spacetimes. *Phys. Rev. Lett.*, 95: 121101, 2005. doi:10.1103/PhysRevLett.95.121101. (cit. on p. 47).

[215] Sebastiano Bernuzzi and Tim Dietrich. Gravitational waveforms from binary neutron star mergers with high-order weighted-essentially-nonoscillatory schemes in numerical relativity. *Phys. Rev. D*, 94(6):064062, 2016. doi:10.1103/PhysRevD.94.064062. (cit. on p. 47).

[216] Francois Foucart et al. Gravitational waveforms from spectral Einstein code simulations: Neutron star-neutron star and low-mass black hole-neutron star binaries. *Phys. Rev. D*, 99(4):044008, 2019. doi:10.1103/PhysRevD.99.044008. (cit. on p. 47).

[217] Karan Jani, James Healy, James A. Clark, Lionel London, Pablo Laguna, and Deirdre Shoemaker. Georgia Tech Catalog of Gravitational Waveforms. *Class. Quant. Grav.*, 33(20):204001, 2016. doi:10.1088/0264-9381/33/20/204001. (cit. on p. 47).

[218] Abdul H. Mroue et al. Catalog of 174 Binary Black Hole Simulations for Gravitational Wave Astronomy. *Phys. Rev. Lett.*, 111(24):241104, 2013. doi:10.1103/PhysRevLett.111.241104. (cit. on p. 47).

[219] Michael Boyle et al. The SXS Collaboration catalog of binary black hole simulations. *Class. Quant. Grav.*, 36(19):195006, 2019. doi:10.1088/1361-6382/ab34e2. (cit. on p. 47).

[220] A. G. Abac et al. Search for Eccentric Black Hole Coalescences during the Third Observing Run of LIGO and Virgo. 8 2023. doi:10.48550/arXiv.2308.03822. (cit. on pp. 47 and 59).

[221] Kip S. Thorne. Multipole expansions of gravitational radiation. *Rev. Mod. Phys.*, 52:299–339, Apr 1980. doi:10.1103/RevModPhys.52.299. URL `https://link.aps.org/doi/10.1103/RevModPhys.52.299`. (cit. on pp. 47 and 48).

[222] Lawrence E. Kidder. Using full information when computing modes of post-Newtonian waveforms from inspiralling compact binaries in circular orbit. *Phys. Rev. D*, 77:044016, 2008. doi:10.1103/PhysRevD.77.044016. (cit. on p. 47).

[223] 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. doi:10.1063/1.1705135. (cit. on p. 48).

[224] P. Szekeres. The Gravitational compass. *J. Math. Phys.*, 6:1387–1391, 1965. doi:10.1063/1.1704788. (cit. on p. 48).

[225] Alessandra Buonanno, Gregory B. Cook, and Frans Pretorius. Inspiral, merger and ring-down of equal-mass black-hole binaries. *Phys. Rev. D*, 75:124018, 2007. doi:10.1103/PhysRevD.75.124018. (cit. on p. 48).

[226] Emanuele Berti, Vitor Cardoso, and Andrei O. Starinets. Quasinormal modes of black holes and black branes. *Class. Quant. Grav.*, 26:163001, 2009. doi:10.1088/0264-9381/26/16/163001. (cit. on p. 48).

[227] S. W. Hawking. Black holes in general relativity. *Commun. Math. Phys.*, 25:152–166, 1972. doi:10.1007/BF01877517. (cit. on p. 48).

[228] Saul A. Teukolsky. Perturbations of a rotating black hole. 1. Fundamental equations for gravitational electromagnetic and neutrino field perturbations. *Astrophys. J.*, 185:635–647, 1973. doi:10.1086/152444. (cit. on p. 48).

[229] William H. Press and Saul A. Teukolsky. Perturbations of a Rotating Black Hole. II. Dynamical Stability of the Kerr Metric. *Astrophys. J.*, 185:649–674, 1973. doi:10.1086/152445. (cit. on p. 48).

[230] Matthew Giesler, Maximiliano Isi, Mark A. Scheel, and Saul Teukolsky. Black Hole Ringdown: The Importance of Overtones. *Phys. Rev. X*, 9(4):041060, 2019. doi:10.1103/PhysRevX.9.041060. (cit. on p. 48).

[231] S. Chandrasekhar and Steven L. Detweiler. The quasi-normal modes of the Schwarzschild black hole. *Proc. Roy. Soc. Lond. A*, 344:441–452, 1975. doi:10.1098/rspa.1975.0112. (cit. on p. 48).

[232] Eric Thrane, Paul D. Lasky, and Yuri Levin. Challenges testing the no-hair theorem with gravitational waves. *Phys. Rev. D*, 96(10):102004, 2017. doi:10.1103/PhysRevD.96.102004. (cit. on p. 48).

[233] Gregorio Carullo et al. Empirical tests of the black hole no-hair conjecture using gravitational-wave observations. *Phys. Rev. D*, 98(10):104020, 2018. doi:10.1103/PhysRevD.98.104020. (cit. on p. 48).

[234] Vishal Baibhav, Emanuele Berti, Vitor Cardoso, and Gaurav Khanna. Black Hole Spectroscopy: Systematic Errors and Ringdown Energy Estimates. *Phys. Rev. D*, 97(4):044048, 2018. doi:10.1103/PhysRevD.97.044048. (cit. on p. 48).

[235] Werner Israel. Event horizons in static vacuum space-times. *Phys. Rev.*, 164:1776–1779, 1967. doi:10.1103/PhysRev.164.1776. (cit. on p. 48).

[236] B. Carter. Axisymmetric Black Hole Has Only Two Degrees of Freedom. *Phys. Rev. Lett.*, 26:331–333, 1971. doi:10.1103/PhysRevLett.26.331. (cit. on p. 48).

[237] Sascha Husa, Sebastian Khan, Mark Hannam, Michael Pürrer, 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. doi:10.1103/PhysRevD.93.044006. (cit. on pp. 49 and 80).

[238] A. Buonanno and T. Damour. Effective one-body approach to general relativistic two-body dynamics. *Phys. Rev. D*, 59:084006, 1999. doi:10.1103/PhysRevD.59.084006. (cit. on p. 49).

[239] Alessandra Buonanno and Thibault Damour. Transition from inspiral to plunge in binary black hole coalescences. *Phys. Rev. D*, 62:064015, 2000. doi:10.1103/PhysRevD.62.064015. (cit. on p. 49).

[240] Thibault Damour and Alessandro Nagar. The Effective One Body description of the Two-Body problem. *Fundam. Theor. Phys.*, 162:211–252, 2011. doi:10.1007/978-90-481-3015-3_7. (cit. on p. 49).

[241] Yi Pan, Alessandra Buonanno, Michael Boyle, Luisa T. Buchman, Lawrence E. Kidder, Harald P. Pfeiffer, and Mark A. Scheel. Inspiral-merger-ringdown multipolar waveforms of nonspinning black-hole binaries using the effective-one-body formalism. *Phys. Rev. D*, 84:124052, 2011. doi:10.1103/PhysRevD.84.124052. (cit. on p. 49).

[242] Antoni Ramos-Buades, Alessandra Buonanno, Héctor Estellés, Mohammed Khalil, Deyan P. Mihaylov, Serguei Ossokine, Lorenzo Pompili, and Mahlet Shiferaw. Next generation of accurate and efficient multipolar precessing-spin effective-one-body waveforms for binary black holes. *Phys. Rev. D*, 108(12):124037, 2023. doi:10.1103/PhysRevD.108.124037. (cit. on pp. 49, 51, 52, and 54).

[243] Deyan P. Mihaylov, Serguei Ossokine, Alessandra Buonanno, Hector Estelles, Lorenzo Pompili, Michael Pürrer, and Antoni Ramos-Buades. pySEOBNR: a software package for the next generation of effective-one-body multipolar waveform models. 3 2023. doi:10.48550/arXiv.2303.18203. (cit. on p. 49).

[244] Andrea Taracchini, Yi Pan, Alessandra Buonanno, Enrico Barausse, Michael Boyle, Tony Chu, Geoffrey Lovelace, Harald P. Pfeiffer, and Mark A. Scheel. Prototype effective-one-body model for nonprecessing spinning inspiral-merger-ringdown waveforms. *Phys. Rev. D*, 86:024011, 2012. doi:10.1103/PhysRevD.86.024011. (cit. on p. 49).

[245] Andrea Taracchini et al. Effective-one-body model for black-hole binaries with generic mass ratios and spins. *Phys. Rev. D*, 89(6):061502, 2014. doi:10.1103/PhysRevD.89.061502. (cit. on p. 49).

[246] Stanislav Babak, Andrea Taracchini, and Alessandra Buonanno. Validating the effective-one-body model of spinning, precessing binary black holes against numerical relativity. *Phys. Rev. D*, 95(2):024010, 2017. doi:10.1103/PhysRevD.95.024010. (cit. on p. 49).

[247] Alejandro Bohé et al. Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors. *Phys. Rev. D*, 95(4):044028, 2017. doi:10.1103/PhysRevD.95.044028. (cit. on p. 49).

[248] Serguei Ossokine et al. Multipolar Effective-One-Body Waveforms for Precessing Binary Black Holes: Construction and Validation. *Phys. Rev. D*, 102(4):044055, 2020. doi:10.1103/PhysRevD.102.044055. (cit. on p. 49).

[249] Antoni Ramos-Buades, Alessandra Buonanno, Mohammed Khalil, and Serguei Ossokine. Effective-one-body multipolar waveforms for eccentric binary black holes with nonprecessing spins. *Phys. Rev. D*, 105(4):044035, 2022. doi:10.1103/PhysRevD.105.044035. (cit. on pp. 49 and 54).

[250] Parameswaran Ajith et al. Phenomenological template family for black-hole coalescence waveforms. *Class. Quant. Grav.*, 24:S689–S700, 2007. doi:10.1088/0264-9381/24/19/S31. (cit. on p. 49).

[251] 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. *Phys. Rev. D*, 93(4):044007, 2016. doi:10.1103/PhysRevD.93.044007. (cit. on pp. 49, 80, and 82).

[252] Sebastian Khan, Katerina Chatziioannou, Mark Hannam, and Frank Ohme. Phenomenological model for the gravitational-wave signal from precessing binary black holes with two-spin effects. *Phys. Rev. D*, 100(2):024059, 2019. doi:10.1103/PhysRevD.100.024059. (cit. on p. 49).

[253] Geraint Pratten, Sascha Husa, Cecilio Garcia-Quiros, Marta Colleoni, Antoni Ramos-Buades, Hector Estelles, 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. *Phys. Rev. D*, 102(6):064001, 2020. doi:10.1103/PhysRevD.102.064001. (cit. on p. 49).

[254] Geraint Pratten et al. Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes. *Phys. Rev. D*, 103(10): 104056, 2021. doi:10.1103/PhysRevD.103.104056. (cit. on pp. 49, 51, and 54).

[255] Jonathan E. Thompson, Eleanor Hamilton, Lionel London, Shrobana Ghosh, Panagiota Kolitsidou, Charlie Hoy, and Mark Hannam. PhenomXO4a: a phenomenological gravitational-wave model for precessing black-hole binaries with higher multipoles and asymmetries. *Phys. Rev. D*, 109(6):063012, 2024. doi:10.1103/PhysRevD.109.063012. (cit. on p. 49).

[256] Tim Dietrich et al. Matter imprints in waveform models for neutron star binaries: Tidal and self-spin effects. *Phys. Rev. D*, 99(2):024029, 2019. doi:10.1103/PhysRevD.99.024029. (cit. on pp. 49 and 85).

[257] Scott E. Field, Chad R. Galley, Jan S. Hesthaven, Jason Kaye, and Manuel Tiglio. Fast prediction and evaluation of gravitational waveforms using surrogate models. *Phys. Rev. X*, 4(3):031006, 2014. doi:10.1103/PhysRevX.4.031006. (cit. on p. 49).

[258] Michael Pürrer. Frequency domain reduced order models for gravitational waves from aligned-spin compact binaries. *Class. Quant. Grav.*, 31(19):195010, 2014. doi:10.1088/0264-9381/31/19/195010. (cit. on pp. 49, 50, and 147).

[259] Michael Pürrer. Frequency domain reduced order model of aligned-spin effective-one-body waveforms with generic mass-ratios and spins. *Phys. Rev. D*, 93(6):064041, 2016. doi:10.1103/PhysRevD.93.064041. (cit. on p. 50).

[260] Roberto Cotesta, Sylvain Marsat, and Michael Pürrer. Frequency domain reduced order model of aligned-spin effective-one-body waveforms with higher-order modes. *Phys. Rev. D*, 101(12):124040, 2020. doi:10.1103/PhysRevD.101.124040. (cit. on p. 50).

[261] Jonathan Blackman, Scott E. Field, Chad R. Galley, Béla Szilágyi, Mark A. Scheel, Manuel Tiglio, and Daniel A. Hemberger. Fast and Accurate Prediction of Numerical Relativity Waveforms from Binary Black Hole Coalescences Using Surrogate Models. *Phys. Rev. Lett.*, 115(12):121102, 2015. doi:10.1103/PhysRevLett.115.121102. (cit. on p. 50).

[262] 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. *Phys. Rev. Research.*, 1:033015, 2019. doi:10.1103/PhysRevResearch.1.033015. (cit. on pp. 50, 52, and 54).

[263] 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. *Phys. Rev. D*, 99(6):064045, 2019. doi:10.1103/PhysRevD.99.064045. (cit. on p. 50).

[264] Tousif Islam, Vijay Varma, Jackie Lodman, Scott E. Field, Gaurav Khanna, Mark A. Scheel, Harald P. Pfeiffer, Davide Gerosa, and Lawrence E. Kidder. Eccentric binary black hole surrogate models for the gravitational waveform and remnant properties: comparable mass, nonspinning case. *Phys. Rev. D*, 103(6):064022, 2021. doi:10.1103/PhysRevD.103.064022. (cit. on pp. 50 and 54).

[265] Lucy M. Thomas, Geraint Pratten, and Patricia Schmidt. Accelerating multimodal gravitational waveforms from precessing compact binaries with artificial neural networks. *Phys. Rev. D*, 106(10):104029, 2022. doi:10.1103/PhysRevD.106.104029. (cit. on pp. 50 and 72).

[266] Christopher J. Moore and Jonathan R. Gair. Novel Method for Incorporating Model Uncertainties into Gravitational Wave Parameter Estimates. *Phys. Rev. Lett.*, 113: 251101, 2014. doi:10.1103/PhysRevLett.113.251101. (cit. on pp. 50, 72, and 94).

[267] Christopher J. Moore, Christopher P. L. Berry, Alvin J. K. Chua, and Jonathan R. Gair. Improving gravitational-wave parameter estimation using Gaussian process regression. *Phys. Rev. D*, 93(6):064001, 2016. doi:10.1103/PhysRevD.93.064001. (cit. on pp. 50 and 72).

[268] Zoheyr Doctor, Ben Farr, Daniel E. Holz, and Michael Pürrer. Statistical Gravitational Waveform Models: What to Simulate Next? *Phys. Rev. D*, 96(12):123011, 2017. doi:10.1103/PhysRevD.96.123011. (cit. on pp. 50 and 72).

[269] Daniel Williams, Ik Siong Heng, Jonathan Gair, James A. Clark, and Bhavesh Khamesra. Precessing numerical relativity waveform surrogate model for binary

black holes: A Gaussian process regression approach. *Phys. Rev. D*, 101(6):063011, 2020. doi:10.1103/PhysRevD.101.063011. (cit. on pp. 50 and 72).

[270] 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, 2012. doi:10.1103/PhysRevD.85.122006. (cit. on pp. 50, 59, 60, and 61).

[271] Michael Pürrer and Carl-Johan Haster. Gravitational waveform accuracy requirements for future ground-based detectors. *Phys. Rev. Res.*, 2(2):023151, 2020. doi:10.1103/PhysRevResearch.2.023151. (cit. on p. 52).

[272] P. Ajith et al. A Template bank for gravitational waveforms from coalescing binary black holes. I. Non-spinning binaries. *Phys. Rev. D*, 77:104017, 2008. doi:10.1103/PhysRevD.77.104017. [Erratum: Phys.Rev.D 79, 129901 (2009)]. (cit. on pp. 52 and 59).

[273] J. Veitch et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Phys. Rev. D*, 91(4):042003, 2015. doi:10.1103/PhysRevD.91.042003. (cit. on pp. 52, 63, and 64).

[274] Michael J. Williams, John Veitch, and Chris Messenger. Nested sampling with normalizing flows for gravitational-wave inference. *Phys. Rev. D*, 103(10):103006, 2021. doi:10.1103/PhysRevD.103.103006. (cit. on pp. 52 and 73).

[275] Thomas D. P. Edwards, Kaze W. K. Wong, Kelvin K. H. Lam, Adam Coogan, Daniel Foreman-Mackey, Maximiliano Isi, and Aaron Zimmerman. ripple: Differentiable and Hardware-Accelerated Waveforms for Gravitational Wave Data Analysis. 2 2023. doi:10.48550/arXiv.2302.05329. (cit. on pp. 52 and 72).

[276] Davide Gerosa, Michael Kesden, Ulrich Sperhake, Emanuele Berti, and Richard O'Shaughnessy. Multi-timescale analysis of phase transitions in precessing black-hole binaries. *Phys. Rev. D*, 92:064016, 2015. doi:10.1103/PhysRevD.92.064016. (cit. on p. 52).

[277] Isobel M. Romero-Shaw, Davide Gerosa, and Nicholas Loutrel. Eccentricity or spin precession? Distinguishing subdominant effects in gravitational-wave data. *Mon. Not. Roy. Astron. Soc.*, 519(4):5352–5357, 2023. doi:10.1093/mnras/stad031. (cit. on p. 52).

[278] Peter T. H. Pang, Juan Calderón Bustillo, Yifan Wang, and Tjonnie G. F. Li. Potential observations of false deviations from general relativity in gravitational wave signals from binary black holes. *Phys. Rev. D*, 98(2):024019, 2018. doi:10.1103/PhysRevD.98.024019. (cit. on pp. 52 and 94).

[279] Pankaj Saini, Marc Favata, and K. G. Arun. Systematic bias on parametrized tests of general relativity due to neglect of orbital eccentricity. *Phys. Rev. D*, 106(8): 084031, 2022. doi:10.1103/PhysRevD.106.084031. (cit. on pp. 52 and 94).

[280] Adam G. M. Lewis, Aaron Zimmerman, and Harald P. Pfeiffer. Fundamental frequencies and resonances from eccentric and precessing binary black hole inspirals. *Class. Quant. Grav.*, 34(12):124001, 2017. doi:10.1088/1361-6382/aa66f4. (cit. on p. 54).

[281] Antoni Ramos-Buades, Sascha Husa, Geraint Pratten, Héctor Estellés, Cecilio García-Quirós, Maite Mateu-Lucena, Marta Colleoni, and Rafel Jaume. First survey of spinning eccentric black hole mergers: Numerical relativity simulations, hybrid waveforms, and parameter estimation. *Phys. Rev. D*, 101(8):083015, 2020. doi:10.1103/PhysRevD.101.083015. (cit. on p. 54).

[282] Xiaolin Liu, Zhoujian Cao, and Zong-Hong Zhu. A higher-multipole gravitational waveform model for an eccentric binary black holes based on the effective-one-body-numerical-relativity formalism. *Class. Quant. Grav.*, 39(3):035009, 2022. doi:10.1088/1361-6382/ac4119. (cit. on p. 54).

[283] Xiaolin Liu, Zhoujian Cao, and Zong-Hong Zhu. Effective-One-Body Numerical-Relativity waveform model for Eccentric spin-precessing binary black hole coalescence. 10 2023. doi:10.48550/arXiv.2310.04552. (cit. on p. 54).

[284] Duncan A. Brown. Searching for gravitational radiation from binary black hole MACHOs in the galactic halo, 12 2004. (cit. on p. 56).

[285] Jerzy Neyman and Egon Sharpe Pearson. On the Problem of the Most Efficient Tests of Statistical Hypotheses. *Phil. Trans. Roy. Soc. Lond. A*, 231(694-706):289–337, 1933. doi:10.1098/rsta.1933.0009. (cit. on p. 56).

[286] Jolien D. E. Creighton and Warren G. Anderson. *Gravitational-Wave Physics and Astronomy: An Introduction to Theory, Experiment and Data Analysis*. 2011. ISBN 9783527408863, 9783527636037. (cit. on p. 58).

[287] Shio Sakon et al. Template bank for compact binary mergers in the fourth observing run of Advanced LIGO, Advanced Virgo, and KAGRA. *Phys. Rev. D*, 109(4):044066, 2024. doi:10.1103/PhysRevD.109.044066. (cit. on p. 59).

[288] Stephen Fairhurst, Rhys Green, Charlie Hoy, Mark Hannam, and Alistair Muir. Two-harmonic approximation for gravitational waveforms from precessing binaries. *Phys. Rev. D*, 102(2):024055, 2020. doi:10.1103/PhysRevD.102.024055. (cit. on pp. 59, 109, and 133).

[289] Kipp Cannon et al. Toward Early-Warning Detection of Gravitational Waves from Compact Binary Coalescence. *Astrophys. J.*, 748:136, 2012. doi:10.1088/0004-637X/748/2/136. (cit. on pp. 61, 129, and 133).

[290] D. Macleod, I. W. Harry, and S. Fairhurst. Fully-coherent all-sky search for gravitational-waves from compact binary coalescences. *Phys. Rev. D*, 93(6):064004, 2016. doi:10.1103/PhysRevD.93.064004. (cit. on pp. 61 and 62).

[291] T. Adams, D. Buskulic, V. Germain, G. M. Guidi, F. Marion, M. Montani, B. Mours, F. Piergiovanni, and G. Wang. Low-latency analysis pipeline for compact binary coalescences in the advanced gravitational wave detector era. *Class. Quant. Grav.*, 33(17):175012, 2016. doi:10.1088/0264-9381/33/17/175012. (cit. on pp. 62, 129, and 130).

[292] Cody Messick et al. Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data. *Phys. Rev. D*, 95:042001, Feb 2017. doi:10.1103/PhysRevD.95.042001. URL `https://link.aps.org/doi/10.1103/PhysRevD.95.042001`. (cit. on p. 62).

[293] Tito Dal Canton et al. Implementing a search for aligned-spin neutron star-black hole systems with advanced ground based gravitational wave detectors. *Phys. Rev. D*, 90(8):082004, 2014. doi:10.1103/PhysRevD.90.082004. (cit. on p. 62).

[294] Kipp Cannon, Chad Hanna, and Jacob Peoples. Likelihood-Ratio Ranking Statistic for Compact Binary Coalescence Candidates with Rate Estimation. 4 2015. doi:10.48550/arXiv.1504.04632. (cit. on p. 62).

[295] Shasvath J. Kapadia et al. A self-consistent method to estimate the rate of compact binary coalescences with a Poisson mixture model. *Class. Quant. Grav.*, 37(4): 045007, 2020. doi:10.1088/1361-6382/ab5f2d. (cit. on p. 62).

[296] Will M. Farr, Jonathan R. Gair, Ilya Mandel, and Curt Cutler. Counting and confusion: Bayesian rate estimation with multiple populations. *Phys. Rev. D*, 91: 023005, Jan 2015. doi:10.1103/PhysRevD.91.023005. URL `https://link.aps.org/doi/10.1103/PhysRevD.91.023005`. (cit. on p. 62).

[297] Leo Tsukada et al. Improved ranking statistics of the GstLAL inspiral search for compact binary coalescences. *Phys. Rev. D*, 108(4):043004, 2023. doi:10.1103/PhysRevD.108.043004. (cit. on p. 62).

[298] Alexander H. Nitz, Tito Dal Canton, Derek Davis, and Steven Reyes. Rapid detection of gravitational waves from compact binary mergers with pycbc live. *Phys. Rev. D*, 98:024050, Jul 2018. doi:10.1103/PhysRevD.98.024050. URL https://link.aps.org/doi/10.1103/PhysRevD.98.024050. (cit. on p. 62).

[299] Tito Dal Canton, Alexander H. Nitz, Bhooshan Gadre, Gareth S. Cabourn Davies, Veronica Villa-Ortega, Thomas Dent, Ian Harry, and Liting Xiao. Real-time Search for Compact Binary Mergers in Advanced LIGO and Virgo's Third Observing Run Using PyCBC Live. *Astrophys. J.*, 923(2):254, 2021. doi:10.3847/1538-4357/ac2f9a. (cit. on p. 62).

[300] Qi Chu et al. SPIIR online coherent pipeline to search for gravitational waves from compact binary coalescences. *Phys. Rev. D*, 105(2):024023, 2022. doi:10.1103/PhysRevD.105.024023. (cit. on pp. 62, 117, and 127).

[301] Sushant Sharma Chaudhary et al. Low-latency gravitational wave alert products and their performance in anticipation of the fourth LIGO-Virgo-KAGRA observing run. *Proc. Nat. Acad. Sci.*, 121(18):e2316474121, 2024. doi:10.1073/pnas.2316474121. (cit. on pp. 62 and 117).

[302] Alexander H. Nitz, Collin Capano, Alex B. Nielsen, Steven Reyes, Rebecca White, Duncan A. Brown, and Badri Krishnan. 1-OGC: The First Open Gravitational-wave Catalog of Binary Mergers from Analysis of Public Advanced LIGO Data. *The Astrophysical Journal*, 872(2):195, 2019. ISSN 1538-4357. doi:10.3847/1538-4357/ab0108. URL http://dx.doi.org/10.3847/1538-4357/ab0108. (cit. on p. 62).

[303] Alexander H. Nitz, Thomas Dent, Gareth S. Davies, Sumit Kumar, Collin D. Capano, Ian Harry, Simone Mozzon, Laura Nuttall, Andrew Lundgren, and Márton Tápai. 2-OGC: Open Gravitational-wave Catalog of Binary Mergers from Analysis of Public Advanced LIGO and Virgo Data. *The Astrophysical Journal*, 891(2):123, 2020. ISSN 1538-4357. doi:10.3847/1538-4357/ab733f. URL http://dx.doi.org/10.3847/1538-4357/ab733f. (cit. on p. 62).

[304] Alexander H. Nitz, Collin D. Capano, Sumit Kumar, Yi-Fan Wang, Shilpa Kastha, Marlin Schäfer, Rahul Dhurkunde, and Miriam Cabero. 3-OGC: Catalog of gravitational waves from compact-binary mergers. *The Astrophysical Journal*, 922(1):76,

November 2021. ISSN 0004-637X, 1538-4357. doi:10.3847/1538-4357/ac1c03. URL http://arxiv.org/abs/2105.09151. (cit. on p. 62).

[305] Alexander H. Nitz, Sumit Kumar, Yi-Fan Wang, Shilpa Kastha, Shichao Wu, Marlin Schäfer, Rahul Dhurkunde, and Collin D. Capano. 4-OGC: Catalog of gravitational waves from compact-binary mergers. *arXiv*, December 2021. URL http://arxiv.org/abs/2112.06878. (cit. on p. 62).

[306] Barak Zackay, Liang Dai, Tejaswi Venumadhav, Javier Roulet, and Matias Zaldarriaga. Detecting gravitational waves with disparate detector responses: Two new binary black hole mergers. *Physical Review D*, 104(6), sep 2021. doi:10.1103/physrevd.104.063030. URL https://doi.org/10.1103%2Fphysrevd.104.063030. (cit. on p. 62).

[307] Tejaswi Venumadhav, Barak Zackay, Javier Roulet, Liang Dai, and Matias Zaldarriaga. New binary black hole mergers in the second observing run of Advanced LIGO and Advanced Virgo. *Phys. Rev. D*, 101:083030, Apr 2020. doi:10.1103/PhysRevD.101.083030. URL https://link.aps.org/doi/10.1103/PhysRevD.101.083030. (cit. on p. 62).

[308] Tejaswi Venumadhav, Barak Zackay, Javier Roulet, Liang Dai, and Matias Zaldarriaga. New search pipeline for compact binary mergers: Results for binary black holes in the first observing run of advanced ligo. *Phys. Rev. D*, 100:023011, Jul 2019. doi:10.1103/PhysRevD.100.023011. URL https://link.aps.org/doi/10.1103/PhysRevD.100.023011. (cit. on p. 62).

[309] Barak Zackay, Tejaswi Venumadhav, Liang Dai, Javier Roulet, and Matias Zaldarriaga. Highly spinning and aligned binary black hole merger in the advanced ligo first observing run. *Phys. Rev. D*, 100:023007, Jul 2019. doi:10.1103/PhysRevD.100.023007. URL https://link.aps.org/doi/10.1103/PhysRevD.100.023007. (cit. on p. 62).

[310] B. P. Abbott et al. Observing gravitational-wave transient GW150914 with minimal assumptions. *Phys. Rev. D*, 93(12):122004, 2016. doi:10.1103/PhysRevD.93.122004. [Addendum: Phys.Rev.D 94, 069903 (2016)]. (cit. on p. 62).

[311] Neil J. Cornish and Tyson B. Littenberg. BayesWave: Bayesian Inference for Gravitational Wave Bursts and Instrument Glitches. *Class. Quant. Grav.*, 32(13):135012, 2015. doi:10.1088/0264-9381/32/13/135012. (cit. on p. 62).

[312] S. Klimenko et al. Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors. *Phys. Rev. D*, 93(4):042004, 2016. doi:10.1103/PhysRevD.93.042004. (cit. on p. 62).

[313] S. Klimenko, I. Yakushin, A. Mercer, and Guenakh Mitselmakher. Coherent method for detection of gravitational wave bursts. *Class. Quant. Grav.*, 25:114029, 2008. doi:10.1088/0264-9381/25/11/114029. (cit. on p. 62).

[314] O. Halim et al. The search of higher multipole radiation in gravitational waves from compact binary coalescences by a minimally-modelled pipeline. *J. Phys. Conf. Ser.*, 2156(1):012081, 2021. doi:10.1088/1742-6596/2156/1/012081. (cit. on p. 62).

[315] Ryan Lynch, Salvatore Vitale, Reed Essick, Erik Katsavounidis, and Florent Robinet. Information-theoretic approach to the gravitational-wave burst detection problem. *Phys. Rev. D*, 95(10):104046, 2017. doi:10.1103/PhysRevD.95.104046. (cit. on p. 62).

[316] 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:e010, March 2019. doi:10.1017/pasa.2019.2. (cit. on p. 63).

[317] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 06 1953. ISSN 0021-9606. doi:10.1063/1.1699114. URL https://doi.org/10.1063/1.1699114. (cit. on p. 64).

[318] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 04 1970. ISSN 0006-3444. doi:10.1093/biomet/57.1.97. URL https://doi.org/10.1093/biomet/57.1.97. (cit. on p. 64).

[319] David J. Earl and Michael W. Deem. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics (Incorporating Faraday Transactions)*, 7(23):3910, January 2005. doi:10.1039/B509983H. (cit. on p. 64).

[320] Michael Betancourt. A Conceptual Introduction to Hamiltonian Monte Carlo. *arXiv e-prints*, January 2017. doi:10.48550/arXiv.1701.02434. (cit. on p. 64).

[321] David W. Hogg and Daniel Foreman-Mackey. Data analysis recipes: Using Markov Chain Monte Carlo. *Astrophys. J. Suppl.*, 236(1):11, 2018. doi:10.3847/1538-4365/aab76e. (cit. on p. 64).

[322] John Skilling. Nested Sampling. *AIP Conference Proceedings*, 735(1):395–405, 11 2004. ISSN 0094-243X. doi:10.1063/1.1835238. URL https://doi.org/10.1063/1.1835238. (cit. on p. 64).

[323] Joshua S Speagle. dynesty: a dynamic nested sampling package for estimating bayesian posteriors and evidences. *Monthly Notices of the Royal Astronomical Society*, 493(3):3132–3158, February 2020. ISSN 1365-2966. doi:10.1093/mnras/staa278. URL http://dx.doi.org/10.1093/mnras/staa278. (cit. on p. 65).

[324] Greg Ashton et al. Nested sampling for physical scientists. *Nature*, 2, 2022. doi:10.1038/s43586-022-00121-x. (cit. on p. 65).

[325] Michele Vallisneri. Use and abuse of the Fisher information matrix in the assessment of gravitational-wave parameter-estimation prospects. *Phys. Rev. D*, 77:042001, 2008. doi:10.1103/PhysRevD.77.042001. (cit. on pp. 68 and 161).

[326] Ziming Wang, Chang Liu, Junjie Zhao, and Lijing Shao. Extending the Fisher Information Matrix in Gravitational-wave Data Analysis. *Astrophys. J.*, 932(2):102, 2022. doi:10.3847/1538-4357/ac6b99. (cit. on p. 68).

[327] C. Radhakrishna Rao. *Information and the Accuracy Attainable in the Estimation of Statistical Parameters*, pages 235–247. Springer New York, New York, NY, 1992. ISBN 978-1-4612-0919-5. doi:10.1007/978-1-4612-0919-5_16. (cit. on p. 68).

[328] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88, 2017. ISSN 1361-8415. doi:https://doi.org/10.1016/j.media.2017.07.005. URL https://www.sciencedirect.com/science/article/pii/S1361841517301135. (cit. on p. 68).

[329] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. Wavenet: A generative model for raw audio. 2016. doi:10.48550/arXiv.1609.03499. (cit. on p. 68).

[330] Yingwei Li, Adams Wei Yu, Tianjian Meng, Ben Caine, Jiquan Ngiam, Daiyi Peng, Junyang Shen, Bo Wu, Yifeng Lu, Denny Zhou, Quoc V. Le, Alan Yuille, and Mingxing Tan. Deepfusion: Lidar-camera deep fusion for multi-modal 3d object detection. 2022. doi:10.48550/arXiv.2203.08195. (cit. on p. 68).

[331] Diksha Khurana, Aditya Koli, Kiran Khatter, and Sukhdev Singh. Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, 82(3):3713–3744, July 2022. ISSN 1573-7721. doi:10.1007/s11042-022-13428-4. URL http://dx.doi.org/10.1007/s11042-022-13428-4. (cit. on p. 68).

[332] Tom B. Brown et al. Language models are few-shot learners. 2020. doi:10.48550/arXiv.2005.14165. (cit. on p. 68).

[333] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. 2015. doi:10.48550/arXiv.1512.03385. (cit. on p. 70).

[334] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. doi:10.48550/arXiv.1706.03762. (cit. on pp. 70 and 147).

[335] Sebastian Ruder. An overview of gradient descent optimization algorithms. 2017. doi:10.48550/arXiv.1609.04747. (cit. on p. 71).

[336] Elena Cuoco et al. Enhancing Gravitational-Wave Science with Machine Learning. *Mach. Learn. Sci. Tech.*, 2(1):011002, 2021. doi:10.1088/2632-2153/abb93a. (cit. on p. 71).

[337] Sebastian Khan and Rhys Green. Gravitational-wave surrogate models powered by artificial neural networks. *Phys. Rev. D*, 103(6):064015, 2021. doi:10.1103/PhysRevD.103.064015. (cit. on p. 71).

[338] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018. URL `http://github.com/google/jax`. (cit. on p. 72).

[339] Kaze W. K. Wong, Maximiliano Isi, and Thomas D. P. Edwards. Fast Gravitational-wave Parameter Estimation without Compromises. *Astrophys. J.*, 958(2):129, 2023. doi:10.3847/1538-4357/acf5cd. (cit. on pp. 72 and 73).

[340] J. McGinn, C. Messenger, I. S. Heng, and M. J. Williams. Generalised gravitational wave burst generation with generative adversarial networks. *Class. Quant. Grav.*, 38(15):155005, 2021. doi:10.1088/1361-6382/ac09cc. (cit. on p. 72).

[341] Reed Essick, Patrick Godwin, Chad Hanna, Lindy Blackburn, and Erik Katsavounidis. iDQ: Statistical Inference of Non-Gaussian Noise with Auxiliary Degrees of Freedom in Gravitational-Wave Detectors. 5 2020. (cit. on p. 72).

[342] Melissa Lopez, Vincent Boudart, Kerwin Buijsman, Amit Reza, and Sarah Caudill. Simulating transient noise bursts in LIGO with generative adversarial networks. *Phys. Rev. D*, 106(2):023027, 2022. doi:10.1103/PhysRevD.106.023027. (cit. on p. 72).

[343] Timothy D. Gebhard, Niki Kilbertus, Ian Harry, and Bernhard Schölkopf. Convolutional neural networks: a magic bullet for gravitational-wave detection? *Phys. Rev. D*, 100(6):063015, 2019. doi:10.1103/PhysRevD.100.063015. (cit. on p. 72).

[344] Paraskevi Nousi, Alexandra E. Koloniari, Nikolaos Passalis, Panagiotis Iosif, Nikolaos Stergioulas, and Anastasios Tefas. Deep residual networks for gravitational wave detection. *Phys. Rev. D*, 108(2):024022, 2023. doi:10.1103/PhysRevD.108.024022. (cit. on p. 72).

[345] Marlin B. Schäfer et al. First machine learning gravitational-wave search mock data challenge. *Phys. Rev. D*, 107(2):023021, 2023. doi:10.1103/PhysRevD.107.023021. (cit. on p. 72).

[346] Vasileios Skliris, Michael R. K. Norman, and Patrick J. Sutton. Real-Time Detection of Unmodelled Gravitational-Wave Transients Using Convolutional Neural Networks. 9 2020. doi:10.48550/arXiv.2009.14611. (cit. on p. 72).

[347] Chayan Chatterjee, Linqing Wen, Foivos Diakogiannis, and Kevin Vinsen. Extraction of binary black hole gravitational wave signals from detector data using deep learning. *Phys. Rev. D*, 104(6):064046, 2021. doi:10.1103/PhysRevD.104.064046. (cit. on p. 72).

[348] Marek J. Szczepańczyk et al. Search for gravitational-wave bursts in the third Advanced LIGO-Virgo run with coherent WaveBurst enhanced by machine learning. *Phys. Rev. D*, 107(6):062002, 2023. doi:10.1103/PhysRevD.107.062002. (cit. on p. 72).

[349] P. Astone, P. Cerdá-Durán, I. Di Palma, M. Drago, F. Muciaccia, C. Palomba, and F. Ricci. New method to observe gravitational waves emitted by core collapse supernovae. *Phys. Rev. D*, 98(12):122002, 2018. doi:10.1103/PhysRevD.98.122002. (cit. on p. 72).

[350] Man Leong Chan, Ik Siong Heng, and Chris Messenger. Detection and classification of supernova gravitational wave signals: A deep learning approach. *Phys. Rev. D*, 102(4):043022, 2020. doi:10.1103/PhysRevD.102.043022. (cit. on p. 72).

[351] Christoph Dreissigacker, Rahul Sharma, Chris Messenger, Ruining Zhao, and Reinhard Prix. Deep-Learning Continuous Gravitational Waves. *Phys. Rev. D*, 100(4): 044009, 2019. doi:10.1103/PhysRevD.100.044009. (cit. on p. 72).

[352] Joe Bayley, Chris Messenger, and Graham Woan. Robust machine learning algorithm to search for continuous gravitational waves. *Physical Review D*, 102

(8), October 2020. ISSN 2470-0029. doi:10.1103/physrevd.102.083024. URL http://dx.doi.org/10.1103/PhysRevD.102.083024. (cit. on p. 72).

[353] Hongyu Shen, Daniel George, E. A. Huerta, and Zhizhen Zhao. Denoising Gravitational Waves using Deep Learning with Recurrent Denoising Autoencoders. 11 2017. doi:10.48550/arXiv.1711.09919. (cit. on p. 72).

[354] Wei Wei and E. A. Huerta. Gravitational Wave Denoising of Binary Black Hole Mergers with Deep Learning. *Phys. Lett. B*, 800:135081, 2020. doi:10.1016/j.physletb.2019.135081. (cit. on p. 72).

[355] Miriam Cabero, Ashish Mahabal, and Jess McIver. GWSkyNet: a real-time classifier for public gravitational-wave candidates. *Astrophys. J. Lett.*, 904(1):L9, 2020. doi:10.3847/2041-8213/abc5b5. (cit. on p. 72).

[356] 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 Phys.*, 18(1):112–117, 2022. doi:10.1038/s41567-021-01425-7. (cit. on pp. 72 and 141).

[357] 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(24):241103, 2021. doi:10.1103/PhysRevLett.127.241103. (cit. on pp. 73, 141, 147, and 148).

[358] 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(17):171403, 2023. doi:10.1103/PhysRevLett.130.171403. (cit. on pp. 73 and 141).

[359] Jurriaan Langendorff, Alex Kolmus, Justin Janquart, and Chris Van Den Broeck. Normalizing Flows as an Avenue to Studying Overlapping Gravitational Wave Signals. *Phys. Rev. Lett.*, 130(17):171402, 2023. doi:10.1103/PhysRevLett.130.171402. (cit. on pp. 73 and 147).

[360] Chayan Chatterjee, Manoj Kovalam, Linqing Wen, Damon Beveridge, Foivos Diakogiannis, and Kevin Vinsen. Rapid Localization of Gravitational Wave Sources from Compact Binary Coalescences Using Deep Learning. *Astrophys. J.*, 959(1):42, 2023. doi:10.3847/1538-4357/ad08b7. (cit. on p. 73).

[361] Jordan McGinn, Arunava Mukherjee, Jessica Irwin, Christopher Messenger, Michael J. Williams, and Ik Siong Heng. Rapid neutron star equation of state

inference with Normalising Flows. 3 2024. doi:10.48550/arXiv.2403.17462. (cit. on pp. 73 and 168).

[362] David Ruhe, Kaze Wong, Miles Cranmer, and Patrick Forré. Normalizing Flows for Hierarchical Bayesian Analysis: A Gravitational Wave Population Study. 11 2022. doi:10.48550/arXiv.2211.09008. (cit. on p. 73).

[363] Konstantin Leyde, Stephen R. Green, Alexandre Toubiana, and Jonathan Gair. Gravitational wave populations and cosmology with neural posterior estimation. *Phys. Rev. D*, 109(6):064056, 2024. doi:10.1103/PhysRevD.109.064056. (cit. on p. 73).

[364] Federico Stachurski, Christopher Messenger, and Martin Hendry. Cosmological Inference using Gravitational Waves and Normalising Flows. 10 2023. doi:10.48550/arXiv.2310.13405. (cit. on p. 73).

[365] Stefano Rinaldi, Walter Del Pozzo, Michela Mapelli, Ana Lorenzo-Medina, and Thomas Dent. Evidence of evolution of the black hole mass function with redshift. *Astron. Astrophys.*, 684:A204, 2024. doi:10.1051/0004-6361/202348161. (cit. on p. 73).

[366] Anarya Ray, Ignacio Magaña Hernandez, Siddharth Mohite, Jolien Creighton, and Shasvath Kapadia. Nonparametric Inference of the Population of Compact Binaries from Gravitational-wave Observations Using Binned Gaussian Processes. *Astrophys. J.*, 957(1):37, 2023. doi:10.3847/1538-4357/acf452. (cit. on p. 73).

[367] Bruce Edelman, Ben Farr, and Zoheyr Doctor. Cover Your Basis: Comprehensive Data-driven Characterization of the Binary Black Hole Population. *Astrophys. J.*, 946(1):16, 2023. doi:10.3847/1538-4357/acb5ed. (cit. on p. 73).

[368] Curt Cutler and Michele Vallisneri. LISA detections of massive black hole inspirals: Parameter extraction errors due to inaccurate template waveforms. *Physical Review D*, 76(10):104018, November 2007. ISSN 1550-7998, 1550-2368. doi:10.1103/PhysRevD.76.104018. URL http://arxiv.org/abs/0707.2982. (cit. on pp. 74, 78, and 106).

[369] Michele Vallisneri and Nicolas Yunes. Stealth Bias in Gravitational-Wave Parameter Estimation. *Physical Review D*, 87(10):102002, May 2013. ISSN 1550-7998, 1550-2368. doi:10.1103/PhysRevD.87.102002. URL http://arxiv.org/abs/1301.2627. (cit. on pp. 74 and 92).

[370] Salvatore Vitale and Walter Del Pozzo. How serious can the stealth bias be in gravitational wave parameter estimation? *Physical Review D*, 89(2):022002, January

2014. ISSN 1550-7998, 1550-2368. doi:10.1103/PhysRevD.89.022002. URL http://arxiv.org/abs/1311.2057. (cit. on pp. 74 and 92).

[371] Andrea Antonelli, Ollie Burke, and Jonathan R. Gair. Noisy neighbours: Inference biases from overlapping gravitational-wave signals. *Monthly Notices of the Royal Astronomical Society*, 507(4):5069–5086, September 2021. ISSN 0035-8711, 1365-2966. doi:10.1093/mnras/stab2358. URL http://arxiv.org/abs/2104.01897. (cit. on pp. 74, 78, 79, 80, and 83).

[372] B. P. Abbott et al. GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence. *Phys. Rev. Lett.*, 119(14):141101, 2017. doi:10.1103/PhysRevLett.119.141101. (cit. on p. 78).

[373] Benjamin P. Abbott et al. GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2. *Phys. Rev. Lett.*, 118(22):221101, 2017. doi:10.1103/PhysRevLett.118.221101. [Erratum: Phys.Rev.Lett. 121, 129901 (2018)]. (cit. on p. 78).

[374] Michele Maggiore and et al. Science case for the Einstein telescope. *Journal of Cosmology and Astroparticle Physics*, 2020(03):050–050, 2020. ISSN 1475-7516. doi:10.1088/1475-7516/2020/03/050. (cit. on p. 78).

[375] Yoshiaki Himemoto, Atsushi Nishizawa, and Atsushi Taruya. Impacts of overlapping gravitational-wave signals on the parameter estimation: Toward the search for cosmological backgrounds. *Physical Review D*, 104(4):044010, August 2021. ISSN 2470-0010, 2470-0029. doi:10.1103/PhysRevD.104.044010. URL http://arxiv.org/abs/2103.14816. (cit. on pp. 78, 83, and 131).

[376] 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):044003, August 2021. ISSN 2470-0010, 2470-0029. doi:10.1103/PhysRevD.104.044003. URL http://arxiv.org/abs/2102.07544. (cit. on pp. 78, 83, and 131).

[377] Philip Relton and Vivien Raymond. Parameter Estimation Bias From Overlapping Binary Black Hole Events In Second Generation Interferometers. *Physical Review D*, 104(8):084039, October 2021. ISSN 2470-0010, 2470-0029. doi:10.1103/PhysRevD.104.084039. URL http://arxiv.org/abs/2103.16225. (cit. on pp. 78, 83, and 131).

[378] Masamune Oguri. Effect of gravitational lensing on the distribution of gravitational waves from distant binary black hole mergers. *Monthly Notices of the Royal Astronomical Society*, 480(3):3842–3855, November 2018. ISSN 0035-8711, 1365-2966.

doi:10.1093/mnras/sty2145. URL http://arxiv.org/abs/1807.02584. (cit. on pp. 78 and 85).

[379] Jeroen Meidam and et al. Parameterized tests of the strong-field dynamics of general relativity using gravitational wave signals from coalescing binary black holes: Fast likelihood calculations and sensitivity of the method. *Physical Review D*, 97(4):044033, February 2018. ISSN 2470-0010, 2470-0029. doi:10.1103/PhysRevD.97.044033. URL http://arxiv.org/abs/1712.08772. (cit. on pp. 78 and 81).

[380] A. R. Williamson, J. Lange, R. O'Shaughnessy, J. A. Clark, Prayush Kumar, J. Calderón Bustillo, and J. Veitch. Systematic challenges for future gravitational wave measurements of precessing binary black holes. *Phys. Rev. D*, 96(12):124041, 2017. doi:10.1103/PhysRevD.96.124041. (cit. on p. 78).

[381] Rossella Gamba, Matteo Breschi, Sebastiano Bernuzzi, Michalis Agathos, and Alessandro Nagar. Waveform systematics in the gravitational-wave inference of tidal parameters and equation of state from binary neutron-star signals. *Phys. Rev. D*, 103:124015, Jun 2021. doi:10.1103/PhysRevD.103.124015. URL https://link.aps.org/doi/10.1103/PhysRevD.103.124015. (cit. on p. 78).

[382] Michael Pürrer and Carl Johan Haster. Gravitational waveform accuracy requirements for future ground-based detectors. *Physical Review Research*, 2(2):1–30, 2020. ISSN 26431564. doi:10.1103/PhysRevResearch.2.023151. (cit. on pp. 78, 82, 106, 113, and 115).

[383] Elia Pizzati, Surabhi Sachdev, Anuradha Gupta, and Bangalore Sathyaprakash. Toward inference of overlapping gravitational-wave signals. *Phys. Rev. D*, 105(10): 104016, 2022. doi:10.1103/PhysRevD.105.104016. (cit. on p. 78).

[384] 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. doi:10.1093/mnras/stad1542. (cit. on pp. 78 and 79).

[385] Luca Reali, Andrea Antonelli, Roberto Cotesta, Ssohrab Borhanian, Mesut Çalışkan, Emanuele Berti, and B. S. Sathyaprakash. The impact of confusion noise on golden binary neutron-star events in next-generation terrestrial observatories. 9 2022. doi:10.48550/arXiv.2209.13452. (cit. on pp. 78, 85, and 94).

[386] Nina Kunert, Peter T. H. Pang, Ingo Tews, Michael W. Coughlin, and Tim Dietrich. Quantifying modeling uncertainties when combining multiple gravitational-wave detections from binary neutron star sources. *Phys. Rev. D*, 105:L061301, Mar 2022.

doi:10.1103/PhysRevD.105.L061301. URL https://link.aps.org/doi/10.1103/PhysRevD.105.L061301. (cit. on pp. 78 and 94).

[387] Christopher J. Moore, Eliot Finch, Riccardo Buscicchio, and Davide Gerosa. Testing general relativity with gravitational-wave catalogs: The insidious nature of waveform systematics. *iScience*, 24(6):102577, June 2021. ISSN 25890042. doi:10.1016/j.isci.2021.102577. URL http://arxiv.org/abs/2103.16486. (cit. on pp. 78, 87, and 88).

[388] Chandra Kant Mishra, K. G. Arun, Bala R. Iyer, and B. S. Sathyaprakash. Parametrized tests of post-Newtonian theory using Advanced LIGO and Einstein Telescope. *Phys. Rev. D*, 82:064010, 2010. doi:10.1103/PhysRevD.82.064010. (cit. on p. 80).

[389] T. G. F. Li, W. Del Pozzo, S. Vitale, C. Van Den Broeck, M. Agathos, J. Veitch, K. Grover, T. Sidery, R. Sturani, and A. Vecchio. Towards a generic test of the strong field dynamics of general relativity using compact binary coalescence. *Phys. Rev. D*, 85:082003, 2012. doi:10.1103/PhysRevD.85.082003. (cit. on p. 80).

[390] Laura Sampson, Neil Cornish, and Nicolas Yunes. Gravitational Wave Tests of Strong Field General Relativity with Binary Inspirals: Realistic Injections and Optimal Model Selection. *Phys. Rev. D*, 87(10):102001, 2013. doi:10.1103/PhysRevD.87.102001. (cit. on p. 81).

[391] Emanuele Berti, Alessandra Buonanno, and Clifford M. Will. Estimating spinning binary parameters and testing alternative theories of gravity with LISA. *Phys. Rev. D*, 71:084025, 2005. doi:10.1103/PhysRevD.71.084025. (cit. on p. 81).

[392] Geraint Pratten et al. Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes. *Physical Review D*, 103(10):104056, May 2021. ISSN 2470-0010, 2470-0029. doi:10.1103/PhysRevD.103.104056. URL https://link.aps.org/doi/10.1103/PhysRevD.103.104056. (cit. on p. 82).

[393] Serguei Ossokine et al. Multipolar effective-one-body waveforms for precessing binary black holes: Construction and validation. *Physical Review D*, 102(4):044055, August 2020. ISSN 2470-0010, 2470-0029. doi:10.1103/PhysRevD.102.044055. URL https://link.aps.org/doi/10.1103/PhysRevD.102.044055. (cit. on p. 82).

[394] P. A. R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.*, 594:A13, 2016. doi:10.1051/0004-6361/201525830. (cit. on p. 85).

[395] Shichao Wu and Alexander H. Nitz. Mock data study for next-generation ground-based detectors: The performance loss of matched filtering due to correlated confusion noise. *Phys. Rev. D*, 107(6):063022, 2023. doi:10.1103/PhysRevD.107.063022. (cit. on p. 85).

[396] Aaron Zimmerman, Carl-Johan Haster, and Katerina Chatziioannou. On combining information from multiple gravitational wave sources. *Physical Review D*, 99(12): 124044, June 2019. ISSN 2470-0010, 2470-0029. doi:10.1103/PhysRevD.99.124044. URL http://arxiv.org/abs/1903.11008. (cit. on p. 86).

[397] Maximiliano Isi, Katerina Chatziioannou, and Will M. Farr. Hierarchical test of general relativity with gravitational waves. *Physical Review Letters*, 123(12):121101, September 2019. ISSN 0031-9007, 1079-7114. doi:10.1103/PhysRevLett.123.121101. URL http://arxiv.org/abs/1904.08011. (cit. on p. 86).

[398] Neil Cornish, Laura Sampson, Nicolas Yunes, and Frans Pretorius. Gravitational Wave Tests of General Relativity with the Parameterized Post-Einsteinian Framework. *Phys. Rev. D*, 84:062003, 2011. doi:10.1103/PhysRevD.84.062003. (cit. on p. 92).

[399] Ling Sun et al. Characterization of systematic error in Advanced LIGO calibration. *Class. Quant. Grav.*, 37(22):225008, 2020. doi:10.1088/1361-6382/abb14e. (cit. on pp. 94 and 101).

[400] Evan D. Hall, Craig Cahillane, Kiwamu Izumi, Rory J. E. Smith, and Rana X. Adhikari. Systematic calibration error requirements for gravitational-wave detectors via the Cramér–Rao bound. *Class. Quant. Grav.*, 36(20):205006, 2019. doi:10.1088/1361-6382/ab368c. (cit. on p. 94).

[401] Jade Powell. Parameter Estimation and Model Selection of Gravitational Wave Signals Contaminated by Transient Detector Noise Glitches. *Class. Quant. Grav.*, 35(15):155017, 2018. doi:10.1088/1361-6382/aacf18. (cit. on p. 94).

[402] Chris Pankow et al. Mitigation of the instrumental noise transient in gravitational-wave data surrounding GW170817. *Phys. Rev. D*, 98(8):084016, 2018. doi:10.1103/PhysRevD.98.084016. (cit. on p. 94).

[403] Anuradha Gupta et al. Possible Causes of False General Relativity Violations in Gravitational Wave Observations. 5 2024. doi:10.48550/arXiv.2405.02197. (cit. on pp. 94 and 170).

[404] Guillaume Dideron, Suvodip Mukherjee, and Luis Lehner. New framework to study unmodeled physics from gravitational wave data. *Phys. Rev. D*, 107(10):104023, 2023. doi:10.1103/PhysRevD.107.104023. (cit. on p. 94).

[405] Lee Lindblom, Benjamin J. Owen, and Duncan A. Brown. Model waveform accuracy standards for gravitational wave data analysis. *Physical Review D*, 78(12):1–12, 2008. ISSN 15507998. doi:10.1103/PhysRevD.78.124020. (cit. on pp. 96, 97, and 113).

[406] L. Santamaria et al. Matching post-Newtonian and numerical relativity waveforms: systematic errors and a new phenomenological model for non-precessing black hole binaries. *Phys. Rev. D*, 82:064016, 2010. doi:10.1103/PhysRevD.82.064016. (cit. on p. 97).

[407] Thibault Damour, Alessandro Nagar, and Miguel Trias. Accuracy and effectualness of closed-form, frequency-domain waveforms for non-spinning black hole binaries. *Phys. Rev. D*, 83:024006, 2011. doi:10.1103/PhysRevD.83.024006. (cit. on p. 97).

[408] Sean T. McWilliams, Bernard J. Kelly, and John G. Baker. Observing mergers of non-spinning black-hole binaries. *Phys. Rev. D*, 82:024014, 2010. doi:10.1103/PhysRevD.82.024014. (cit. on p. 100).

[409] R. Abbott et al. Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences. *Astrophys. J. Lett.*, 915(1):L5, 2021. doi:10.3847/2041-8213/ac082e. (cit. on p. 101).

[410] A. R. Williamson, J. Lange, R. O'Shaughnessy, J. A. Clark, Prayush Kumar, J. Calderón Bustillo, and J. Veitch. Systematic challenges for future gravitational wave measurements of precessing binary black holes. *Physical Review D*, 96(12): 124041, December 2017. doi:10.1103/PhysRevD.96.124041. (cit. on pp. 106 and 109).

[411] Jacob Lange, Richard O'Shaughnessy, and Monica Rizzo. Rapid and accurate parameter inference for coalescing, precessing compact binaries, May 2018. URL `http://arxiv.org/abs/1805.10457`. (cit. on p. 106).

[412] Sylvia Biscoveanu, Maximiliano Isi, Vijay Varma, and Salvatore Vitale. Measuring the spins of heavy binary black holes. *Phys. Rev. D*, 104(10):103018, 2021. doi:10.1103/PhysRevD.104.103018. (cit. on pp. 108 and 109).

[413] Vijay Varma and Parameswaran Ajith. Effects of nonquadrupole modes in the detection and parameter estimation of black hole binaries with nonprecessing spins. *Phys. Rev. D*, 96(12):124024, 2017. doi:10.1103/PhysRevD.96.124024. (cit. on pp. 108 and 109).

[414] Marta Colleoni, Maite Mateu-Lucena, Héctor Estellés, Cecilio García-Quirós, David Keitel, Geraint Pratten, Antoni Ramos-Buades, and Sascha Husa. Towards the routine use of subdominant harmonics in gravitational-wave inference: Reanalysis of GW190412 with generation X waveform models. *Phys. Rev. D*, 103(2):024029, 2021. doi:10.1103/PhysRevD.103.024029. (cit. on pp. 108 and 109).

[415] Salvatore Vitale, Ryan Lynch, John Veitch, Vivien Raymond, and Riccardo Sturani. Measuring the spin of black holes in binary systems using gravitational waves. *Phys. Rev. Lett.*, 112(25):251101, 2014. doi:10.1103/PhysRevLett.112.251101. (cit. on p. 109).

[416] N. V. Krishnendu and Frank Ohme. Interplay of spin-precession and higher harmonics in the parameter estimation of binary black holes. *Phys. Rev. D*, 105(6): 064012, 2022. doi:10.1103/PhysRevD.105.064012. (cit. on p. 109).

[417] B. P. Abbott and et al. Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Reviews in Relativity*, 23(1):3, December 2020. ISSN 2367-3613, 1433-8351. doi:10.1007/s41114-020-00026-9. URL https://link.springer.com/10.1007/s41114-020-00026-9. (cit. on p. 109).

[418] Francesco Pannarale, Emanuele Berti, Koutarou Kyutoku, Benjamin D. Lackey, and Masaru Shibata. Aligned spin neutron star-black hole mergers: a gravitational waveform amplitude model. *Phys. Rev. D*, 92(8):084050, 2015. doi:10.1103/PhysRevD.92.084050. (cit. on p. 110).

[419] Rossella Gamba, Sarp Akçay, Sebastiano Bernuzzi, and Jake Williams. Effective-one-body waveforms for precessing coalescing compact binaries with post-Newtonian Twist, 11 2021. (cit. on p. 110).

[420] Danilo Chiaramello and Alessandro Nagar. Faithful analytical effective-one-body waveform model for spin-aligned, moderately eccentric, coalescing black hole binaries. *Phys. Rev. D*, 101(10):101501, 2020. doi:10.1103/PhysRevD.101.101501. (cit. on p. 110).

[421] Tim Dietrich, Anuradha Samajdar, Sebastian Khan, Nathan K. Johnson-McDaniel, Reetika Dudi, and Wolfgang Tichy. Improving the NRTidal model for binary neutron star systems. *Phys. Rev. D*, 100(4):044003, 2019. doi:10.1103/PhysRevD.100.044003. (cit. on p. 113).

[422] Soumi De, Daniel Finstad, James M. Lattimer, Duncan A. Brown, Edo Berger, and Christopher M. Biwer. Tidal Deformabilities and Radii of Neutron Stars

from the Observation of GW170817. *Phys. Rev. Lett.*, 121(9):091102, 2018. doi:10.1103/PhysRevLett.121.091102. [Erratum: Phys.Rev.Lett. 121, 259902 (2018)]. (cit. on p. 117).

[423] K. P. Mooley, A. T. Deller, O. Gottlieb, E. Nakar, G. Hallinan, S. Bourke, D. A. Frail, A. Horesh, A. Corsi, and K. Hotokezaka. Superluminal motion of a relativistic jet in the neutron-star merger GW170817. *Nature*, 561(7723):355–359, 2018. doi:10.1038/s41586-018-0486-3. (cit. on p. 117).

[424] Collin D. Capano, Ingo Tews, Stephanie M. Brown, Ben Margalit, Soumi De, Sumit Kumar, Duncan A. Brown, Badri Krishnan, and Sanjay Reddy. Stringent constraints on neutron-star radii from multimessenger observations and nuclear theory. *Nature Astron.*, 4(6):625–632, 2020. doi:10.1038/s41550-020-1014-6. (cit. on p. 117).

[425] M. Nicholl et al. The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/VIRGO GW170817. III. Optical and UV Spectra of a Blue Kilonova From Fast Polar Ejecta. *Astrophys. J. Lett.*, 848(2):L18, 2017. doi:10.3847/2041-8213/aa9029. (cit. on p. 117).

[426] Eemeli Annala, Tyler Gorda, Aleksi Kurkela, and Aleksi Vuorinen. Gravitational-wave constraints on the neutron-star-matter Equation of State. *Phys. Rev. Lett.*, 120(17):172703, 2018. doi:10.1103/PhysRevLett.120.172703. (cit. on p. 117).

[427] Daniel Kasen, Brian Metzger, Jennifer Barnes, Eliot Quataert, and Enrico Ramirez-Ruiz. Origin of the heavy elements in binary neutron-star mergers from a gravitational wave event. *Nature*, 551:80, 2017. doi:10.1038/nature24453. (cit. on p. 117).

[428] Ben Margalit and Brian D. Metzger. Constraining the Maximum Mass of Neutron Stars From Multi-Messenger Observations of GW170817. *Astrophys. J. Lett.*, 850 (2):L19, 2017. doi:10.3847/2041-8213/aa991c. (cit. on p. 117).

[429] P. S. Cowperthwaite et al. The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. II. UV, Optical, and Near-infrared Light Curves and Comparison to Kilonova Models. *Astrophys. J. Lett.*, 848(2):L17, 2017. doi:10.3847/2041-8213/aa8fc7. (cit. on p. 117).

[430] M. Soares-Santos et al. The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. I. Discovery of the Optical Counterpart Using the Dark Energy Camera. *Astrophys. J. Lett.*, 848(2):L16, 2017. doi:10.3847/2041-8213/aa9059. (cit. on p. 117).

[431] A. J. Cooper, O. Gupta, Z. Wadiasingh, R. A. M. J. Wijers, O. M. Boersma, I. Andreoni, A. Rowlinson, and K. Gourdji. Pulsar revival in neutron star mergers: multi-messenger prospects for the discovery of pre-merger coherent radio emission. *Monthly Notices of the Royal Astronomical Society*, 10 2022. doi:10.1093/mnras/stac3580. (cit. on p. 117).

[432] Elias R. Most and Alexander A. Philippov. Electromagnetic precursors to gravitational wave events: Numerical simulations of flaring in pre-merger binary neutron star magnetospheres. *Astrophys. J. Lett.*, 893(1):L6, 2020. doi:10.3847/2041-8213/ab8196. (cit. on p. 117).

[433] Brian D. Metzger and Anthony L. Piro. Optical and X-ray emission from stable millisecond magnetars formed from the merger of binary neutron stars. *Mon. Not. Roy. Astron. Soc.*, 439:3916–3930, 2014. doi:10.1093/mnras/stu247. (cit. on p. 117).

[434] David Radice, Albino Perego, Kenta Hotokezaka, Steven A. Fromm, Sebastiano Bernuzzi, and Luke F. Roberts. Binary Neutron Star Mergers: Mass Ejection, Electromagnetic Counterparts and Nucleosynthesis. *Astrophys. J.*, 869(2):130, 2018. doi:10.3847/1538-4357/aaf054. (cit. on p. 117).

[435] Manoj Kovalam, Md Anwarul Kaium Patwary, Anala K. Sreekumar, Linqing Wen, Fiona H. Panther, and Qi Chu. Early Warnings of Binary Neutron Star Coalescence Using the SPIIR Search. *Astrophys. J. Lett.*, 927(1):L9, 2022. doi:10.3847/2041-8213/ac5687. (cit. on p. 117).

[436] Ryan Magee et al. First demonstration of early warning gravitational wave alerts. *Astrophys. J. Lett.*, 910(2):L21, 2021. doi:10.3847/2041-8213/abed54. (cit. on p. 117).

[437] Surabhi Sachdev et al. An Early-warning System for Electromagnetic Follow-up of Gravitational-wave Events. *Astrophys. J. Lett.*, 905(2):L25, 2020. doi:10.3847/2041-8213/abc753. (cit. on pp. 117 and 136).

[438] Alexander H. Nitz, Marlin Schäfer, and Tito Dal Canton. Gravitational-wave Merger Forecasting: Scenarios for the Early Detection and Localization of Compact-binary Mergers with Ground-based Observatories. *Astrophys. J. Lett.*, 902:L29, 2020. doi:10.3847/2041-8213/abbc10. (cit. on p. 117).

[439] Ryan Magee and Ssohrab Borhanian. Observing Scenarios for the Next Decade of Early Warning Detection of Binary Neutron Stars. *Astrophys. J.*, 935(2):139, 2022. doi:10.3847/1538-4357/ac7f33. (cit. on p. 117).

[440] Biswajit Banerjee et al. Pre-merger alert to detect the very-high-energy prompt emission from binary neutron-star mergers: Einstein Telescope and Cherenkov Telescope Array synergy, 12 2022. (cit. on p. 117).

[441] Alexander H. Nitz, Tito Dal Canton, Derek Davis, and Steven Reyes. Rapid detection of gravitational waves from compact binary mergers with PyCBC Live. *Phys. Rev. D*, 98(2):024050, 2018. doi:10.1103/PhysRevD.98.024050. (cit. on p. 117).

[442] Cody Messick et al. Analysis Framework for the Prompt Discovery of Compact Binary Mergers in Gravitational-wave Data. *Phys. Rev. D*, 95(4):042001, 2017. doi:10.1103/PhysRevD.95.042001. (cit. on p. 117).

[443] F. Aubin et al. The MBTA pipeline for detecting compact binary coalescences in the third LIGO–Virgo observing run. *Class. Quant. Grav.*, 38(9):095004, 2021. doi:10.1088/1361-6382/abe913. (cit. on pp. 117 and 129).

[444] Gregory Baltus, Justin Janquart, Melissa Lopez, Amit Reza, Sarah Caudill, and Jean-Rene Cudell. Convolutional neural networks for the detection of the early inspiral of a gravitational-wave signal. *Phys. Rev. D*, 103:102003, 2021. doi:10.1103/PhysRevD.103.102003. (cit. on p. 117).

[445] Wei Wei and E. A. Huerta. Deep learning for gravitational wave forecasting of neutron star mergers. *Phys. Lett. B*, 816:136185, 2021. doi:10.1016/j.physletb.2021.136185. (cit. on p. 117).

[446] Hang Yu, Rana X. Adhikari, Ryan Magee, Surabhi Sachdev, and Yanbei Chen. Early warning of coalescing neutron-star and neutron-star-black-hole binaries from the nonstationary noise background using neural networks. *Phys. Rev. D*, 104(6): 062004, 2021. doi:10.1103/PhysRevD.104.062004. (cit. on p. 117).

[447] Grégory Baltus, Jean-René Cudell, Justin Janquart, Melissa Lopez, Sarah Caudill, and Amit Reza. Detecting the early inspiral of a gravitational-wave signal with convolutional neural networks. In *2021 International Conference on Content-Based Multimedia Indexing (CBMI)*, pages 1–6, 2021. doi:10.1109/CBMI50038.2021.9461919. (cit. on p. 117).

[448] Alexander H. Nitz and Tito Dal Canton. Pre-merger Localization of Compact-binary Mergers with Third-generation Observatories. *Astrophys. J. Lett.*, 917(2):L27, 2021. doi:10.3847/2041-8213/ac1a75. (cit. on p. 117).

[449] Samuele Ronchini, Marica Branchesi, Gor Oganesyan, Biswajit Banerjee, Ulyana Dupletsa, Giancarlo Ghirlanda, Jan Harms, Michela Mapelli, and Filippo Santoliquido. Perspectives for multi-messenger astronomy with the next generation

of gravitational-wave detectors and high-energy satellites. *Astronomy & Astrophysics*, 665:A97, September 2022. ISSN 0004-6361, 1432-0746. doi:10.1051/0004-6361/202243705. URL http://arxiv.org/abs/2204.01746. (cit. on pp. 117 and 136).

[450] Man Leong Chan, Chris Messenger, Ik Siong Heng, and Martin Hendry. Binary Neutron Star Mergers and Third Generation Detectors: Localization and Early Warning. *Physical Review D*, 97(12):123014, June 2018. ISSN 2470-0010, 2470-0029. doi:10.1103/PhysRevD.97.123014. URL http://arxiv.org/abs/1803.09680. (cit. on p. 117).

[451] Sarp Akcay. Forecasting Gamma-Ray Bursts Using Gravitational Waves. *Annalen Phys.*, 531(1):1800365, 2019. doi:10.1002/andp.201800365. (cit. on p. 117).

[452] Leo P. Singer and Larry R. Price. Rapid Bayesian position reconstruction for gravitational-wave transients. *Phys. Rev. D*, 93(2):024013, 2016. doi:10.1103/PhysRevD.93.024013. (cit. on pp. 118, 127, and 130).

[453] J. Veitch, I. Mandel, B. Aylott, B. Farr, V. Raymond, C. Rodriguez, M. van der Sluys, V. Kalogera, and A. Vecchio. Estimating parameters of coalescing compact binaries with proposed advanced detector networks. *Phys. Rev. D*, 85:104045, 2012. doi:10.1103/PhysRevD.85.104045. (cit. on p. 118).

[454] Stephen Fairhurst. Triangulation of gravitational wave sources with a network of detectors. *New J. Phys.*, 11:123006, 2009. doi:10.1088/1367-2630/11/12/123006. [Erratum: New J.Phys. 13, 069602 (2011)]. (cit. on pp. 118 and 133).

[455] Reinhard Prix and Badri Krishnan. Targeted search for continuous gravitational waves: Bayesian versus maximum-likelihood statistics. *Class. Quant. Grav.*, 26: 204013, 2009. doi:10.1088/0264-9381/26/20/204013. (cit. on p. 120).

[456] John T. Whelan, Reinhard Prix, Curt J. Cutler, and Joshua L. Willis. New Coordinates for the Amplitude Parameter Space of Continuous Gravitational Waves. *Class. Quant. Grav.*, 31:065002, 2014. doi:10.1088/0264-9381/31/6/065002. (cit. on p. 120).

[457] Stephen Fairhurst. Localization of transient gravitational wave sources: beyond triangulation. *Class. Quant. Grav.*, 35(10):105002, 2018. doi:10.1088/1361-6382/aab675. (cit. on p. 120).

[458] Takuya Tsutsui, Kipp Cannon, and Leo Tsukada. High speed source localization in searches for gravitational waves from compact object collisions. *Physical Review*

D, 103(4), Feb 2021. ISSN 2470-0029. doi:10.1103/physrevd.103.043011. URL http://dx.doi.org/10.1103/PhysRevD.103.043011. (cit. on p. 125).

[459] K. M. Gorski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelman. Healpix: A framework for high-resolution discretization and fast analysis of data distributed on the sphere. *Astrophys.J.*, 622:759–771, 2005. doi:10.1086/427976. (cit. on p. 127).

[460] Wen Zhao and Linqing Wen. Localization accuracy of compact binary coalescences detected by the third-generation gravitational-wave detectors and implication for cosmology. *Phys. Rev. D*, 97(6):064031, 2018. doi:10.1103/PhysRevD.97.064031. (cit. on p. 128).

[461] Pratyusava Baral, Soichiro Morisaki, Ignacio Magaña Hernandez, and Jolien Creighton. Localization of binary neutron star mergers with a single cosmic explorer. *Phys. Rev. D*, 108(4):043010, 2023. doi:10.1103/PhysRevD.108.043010. (cit. on p. 128).

[462] B. P. Abbott et al. GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4\,M\odot$. *The Astrophysical Journal*, 892(1):L3, 2020. ISSN 2041-8205. doi:10.3847/2041-8213/ab75f5. (cit. on p. 129).

[463] Soichiro Morisaki. Accelerating parameter estimation of gravitational waves from compact binary coalescence using adaptive frequency resolutions. *Phys. Rev. D*, 104 (4):044062, 2021. doi:10.1103/PhysRevD.104.044062. (cit. on pp. 130 and 148).

[464] 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):122006, June 2012. ISSN 1550-7998. doi:10.1103/PhysRevD.85.122006. URL https://link.aps.org/doi/10.1103/PhysRevD.85.122006. (cit. on p. 130).

[465] Elia Pizzati, Surabhi Sachdev, Anuradha Gupta, and Bangalore Sathyaprakash. Towards inference of overlapping gravitational wave signals. *Physical Review D*, 105(10):104016, May 2022. ISSN 2470-0010, 2470-0029. doi:10.1103/PhysRevD.105.104016. URL http://arxiv.org/abs/2102.07692. (cit. on p. 131).

[466] Philip Relton, Andrea Virtuoso, Sophie Bini, Vivien Raymond, Ian Harry, Marco Drago, Claudia Lazzaro, Andrea Miani, and Shubhanshu Tiwari. Addressing the challenges of detecting time-overlapping compact binary coalescences. *Physical Review D*, 106(10):104045, November 2022. ISSN 2470-0010, 2470-0029.

doi:10.1103/PhysRevD.106.104045. URL http://arxiv.org/abs/2208.00261. (cit. on p. 131).

[467] Luc Blanchet, Thibault Damour, Bala R. Iyer, Clifford M. Will, and Alan G. Wiseman. Gravitational radiation damping of compact binary systems to second postNewtonian order. *Phys. Rev. Lett.*, 74:3515–3518, 1995. doi:10.1103/PhysRevLett.74.3515. (cit. on p. 131).

[468] Eric Poisson. Gravitational waves from inspiraling compact binaries: The Quadrupole moment term. *Phys. Rev. D*, 57:5287–5290, 1998. doi:10.1103/PhysRevD.57.5287. (cit. on p. 131).

[469] 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. *Class. Quant. Grav.*, 29:175004, 2012. doi:10.1088/0264-9381/29/17/175004. (cit. on p. 131).

[470] Connor McIsaac, Charlie Hoy, and Ian Harry. A search technique to observe precessing compact binary mergers in the advanced detector era. 3 2023. (cit. on p. 133).

[471] Takuya Tsutsui, Atsushi Nishizawa, and Soichiro Morisaki. Early warning of precessing compact binary merger with third-generation gravitational-wave detectors. *Phys. Rev. D*, 104(6):064013, 2021. doi:10.1103/PhysRevD.104.064013. (cit. on p. 133).

[472] Sizheng Ma, Zhoujian Cao, Chun-Yu Lin, Hsing-Po Pan, and Hwei-Jang Yo. Gravitational wave source localization for eccentric binary coalesce with a ground-based detector network. *Physical Review D*, 96(8):084046, October 2017. ISSN 2470-0010, 2470-0029. doi:10.1103/PhysRevD.96.084046. URL http://arxiv.org/abs/1710.02965. (cit. on p. 133).

[473] Shasvath J. Kapadia, Mukesh Kumar Singh, Md Arif Shaikh, Deep Chatterjee, and Parameswaran Ajith. Of Harbingers and Higher Modes: Improved gravitational-wave early-warning of compact binary mergers. *Astrophys. J. Lett.*, 898(2):L39, 2020. doi:10.3847/2041-8213/aba42d. (cit. on p. 133).

[474] Mukesh Kumar Singh, Shasvath J. Kapadia, Md Arif Shaikh, Deep Chatterjee, and Parameswaran Ajith. Improved early warning of compact binary mergers using higher modes of gravitational radiation: A population study. *Mon. Not. Roy. Astron. Soc.*, 502(2):1612–1622, 2021. doi:10.1093/mnras/stab125. (cit. on p. 133).

[475] Qian Hu and John Veitch. Data release for pre-merger localization of binary neutron stars in third generation gravitational wave detectors, August 2023. URL https://doi.org/10.5281/zenodo.8297806. (cit. on p. 139).

[476] Ivan Kobyzev, Simon JD Prince, and Marcus A Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE transactions on pattern analysis and machine intelligence*, 43(11):3964–3979, 2020. doi:10.1109/TPAMI.2020.2992934. (cit. on p. 142).

[477] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021. doi:10.5555/3546258.3546315. (cit. on p. 142).

[478] Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. 2015. doi:10.48550/arXiv.1410.8516. URL https://arxiv.org/abs/1410.8516. (cit. on p. 143).

[479] Thomas Müller, Brian McWilliams, Fabrice Rousselle, Markus Gross, and Jan Novák. Neural importance sampling. 2019. doi:10.48550/arXiv.1808.03856. URL https://arxiv.org/abs/1808.03856. (cit. on p. 144).

[480] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Cubic-spline flows. 2019. doi:10.48550/arXiv.1906.02145. URL https://arxiv.org/abs/1906.02145. (cit. on p. 144).

[481] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural spline flows. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. doi:10.48550/arXiv.1906.04032. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/7ac71d433f282034e088473244df8c02-Paper.pdf. (cit. on p. 144).

[482] Michael J. Williams, jmcginn, federicostak, and John Veitch. uofgravity/glasflow: v0.4.0, July 2024. URL https://doi.org/10.5281/zenodo.12626838. (cit. on p. 145).

[483] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. nflows: normalizing flows in PyTorch, November 2020. URL https://doi.org/10.5281/zenodo.4296287. (cit. on p. 145).

[484] Danilo Jimenez Rezende and Shakir Mohamed. Variational inference with normalizing flows. 2016. doi:10.48550/arXiv.1505.05770. URL https://arxiv.org/abs/1505.05770. (cit. on p. 146).

[485] Alexey Dosovitskiy et al. An image is worth 16x16 words: Transformers for image recognition at scale, 2021. (cit. on pp. 147 and 148).

[486] Maximilian Dax, Stephen R. Green, Jonathan Gair, Nihar Gupte, Michael Pürrer, Vivien Raymond, Jonas Wildberger, Jakob H. Macke, Alessandra Buonanno, and Bernhard Schölkopf. Real-time gravitational-wave inference for binary neutron stars using machine learning. 7 2024. doi:10.48550/arXiv.2407.09602. (cit. on pp. 148, 152, 154, and 168).

[487] Serena Vinciguerra, John Veitch, and Ilya Mandel. Accelerating gravitational wave parameter estimation with multi-band template interpolation. *Class. Quant. Grav.*, 34(11):115006, 2017. doi:10.1088/1361-6382/aa6d44. (cit. on pp. 148 and 150).

[488] Neil J. Cornish. Fast Fisher Matrices and Lazy Likelihoods. 7 2010. doi:10.48550/arXiv.1007.4820. (cit. on p. 151).

[489] Barak Zackay, Liang Dai, and Tejaswi Venumadhav. Relative Binning and Fast Likelihood Evaluation for Gravitational Wave Parameter Estimation. 6 2018. doi:10.48550/arXiv.1806.08792. (cit. on p. 151).

[490] Nathaniel Leslie, Liang Dai, and Geraint Pratten. Mode-by-mode relative binning: Fast likelihood estimation for gravitational waveforms with spin-orbit precession and multiple harmonics. *Phys. Rev. D*, 104(12):123030, 2021. doi:10.1103/PhysRevD.104.123030. (cit. on p. 151).

[491] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 53(2):217–288, 2011. doi:10.48550/arXiv.0909.4061. (cit. on p. 152).

[492] Maximilian Dax, Stephen R. Green, Jonathan Gair, Michael Deistler, Bernhard Schölkopf, and Jakob H. Macke. Group equivariant neural posterior estimation. 11 2021. doi:10.48550/arXiv.2111.13139. (cit. on p. 155).

[493] Patrick J. Sutton et al. X-Pipeline: An Analysis package for autonomous gravitational-wave burst searches. *New J. Phys.*, 12:053034, 2010. doi:10.1088/1367-2630/12/5/053034. (cit. on p. 155).

[494] Sean E. Urban and P. Kenneth Seidelmann. *Explanatory Supplement to the Astronomical Almanac*. University Science Books, 3 edition, 2013. ISBN 978-1-891389-85-6, 978-1-938787-54-6. URL https://uscibooks.aip.org/books/explanatory-supplement-to-the-astronomical-almanac-3rd-edition/. (cit. on p. 155).

[495] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. doi:10.48550/arXiv.1412.6980. (cit. on p. 159).