Master's Thesis

Applications of convolutional neural network and Hilbert-Huang transform to gravitational-wave data analysis

Seiya Sasaoka

A thesis submitted for the degree of Master of Science

in the

Department of Physics

Tokyo Institute of Technology

February 5, 2024

Abstract

The era of gravitational-wave physics began with the first detection of a gravitational-wave signal from a binary black hole merger in 2015. Ground-based interferometric detectors now regularly observe gravitational waves from compact binary coalescences, with over 100 detections to date. These observations have had a significant impact on various areas of physics, such as astrophysics, cosmology, and nuclear physics. As detector sensitivity improves and new detectors are constructed, the frequency of detections is expected to increase further. This highlights the growing need for computationally efficient algorithms to detect these signals and localize their sources. Reducing computational costs is also important to enable follow-up observations by electromagnetic telescopes. To address this demand, machine learning has emerged as a potential solution and has been actively studied over the past decade. Although machine learning-based methods have shown promising performance in various tasks, understanding their predictions can be challenging due to the intricate nature of the models, which are characterized by an extremely large number of parameters. In the first two studies presented in this thesis, we address this challenge by applying explainable artificial intelligence techniques to improve the reliability of gravitational-wave detection models. Our visualization of the rationale behind the models' predictions provides interpretability and offers insights to enhance their efficiency.

Core-collapse supernovae represent another anticipated source of gravitational waves by current and future ground-based detectors. Detecting these events is crucial for gaining insights into the explosion mechanism and the internal structure of neutron stars. To infer physical information from detected gravitational-wave signals, accurate extraction of the frequencies is essential. The Hilbert-Huang transform is a viable option for this purpose, as it is not affected by the trade-off relationship between temporal duration and bandwidth and does not assume any specific waveform morphologies. In the final research covered in this thesis, we employ this technique to analyze gravitational waves originating from core-collapse supernovae. We present a study toward extracting the frequencies of gravitational-wave modes arising from proto-neutron star oscillations to estimate physical properties of these stars. Our approach demonstrates comparable accuracy to the conventional short-time Fourier transform-based method, indicating its potential as a new means of inferring physical parameters of proto-neutron stars from gravitational-wave signals.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor, Kentaro Somiya, for his continuous support throughout my research journey. His broad knowledge in the field of gravitational waves and his constructive feedback have greatly improved my research. I am also grateful to him for providing opportunities for me to participate in various domestic and international conferences, which have further enriched my academic experience.

I would like to extend my sincere gratitude to my adviser, Hirotaka Takahashi, for his invaluable guidance and support throughout my research. The discussions with him in our regular meetings have always guided my study in the right direction. I also appreciate the computational resources and code he shared, which enabled me to perform computations seamlessly.

Special thanks go to those who contributed to my published papers. Diego Dominguez played an instrumental role in proofreading and refining the English of my papers. His academic writing skills were crucial in ensuring their clarity. I had many fruitful discussions about coding and analyses with Naoki Koyama, Suyog Garg, and Yilun Hou. Yuto Omae and Yusuke Sakai also provided insightful questions and comments. I am thankful to the LIGO-Virgo-KAGRA supernova working group and the CBC working group for constructive comments on my presentations.

This thesis makes use of various software tools and observation data from the LIGO-Virgo-KAGRA collaboration. I express my gratitude to the numerous collaborators involved. In particular, I would like to express my appreciation to Jade Powell for sharing their supernova simulation data. I am also thankful to Ken-ichi Oohara for developing parallelized C code for the Hilbert-Huang transform.

I express my heartfelt appreciation to all the other members I have worked with in our laboratory. I am thankful to Lecturer Ken-ichi Harada for guiding me through experiments on interferometers. I extend my thanks to former students, Sotatsu Otabe, Makoto Kuribayashi, Koki Tachihara, Takanori Suzuki, Sapna Hassanaly, and Hayato Tanaka, as well as current students, Homare Abe, Kaido Suzuki, Sakino Takeda, Kotaro Takeguchi, Wataru Usukura, Daiki Haba, Junlang Li, Ryo Iden, and Kento Takeshita for enriching my research life.

Finally, I express my gratitude to my friends and my family.

List of Publications

Below is a list of publications I was involved in during my master's program.

[1] Seiya Sasaoka, Yilun Hou, Kentaro Somiya, and Hirotaka Takahashi, Localization of gravitational waves using machine learning, Phys. Rev. D **105**, 103030 (2022).

[2] Seiya Sasaoka, Naoki Koyama, Diego Dominguez, Yusuke Sakai, Kentaro Somiya, Yuto Omae, and Hirotaka Takahashi, Visualizing convolutional neural network for classifying gravitational waves from core-collapse supernovae, Phys. Rev. D **108**, 123033 (2023).

[3] Seiya Sasaoka, Naoki Koyama, Diego Dominguez, Yusuke Sakai, Kentaro Somiya, Yuto Omae, and Hirotaka Takahashi, Comparative study of 1D and 2D convolutional neural network models with attribution analysis for gravitational wave detection from compact binary coalescences, Phys. Rev. D **109**, 043011 (2024).

Chapters 5 and 6 of this thesis are reprinted from papers [2] and [3], respectively. The paper [1] is based on my Bachelor's thesis and is not included in this thesis.

Contents

List of Figures

List of Tables

Glossary

Notation

Constants. The speed of light is denoted by c, the gravitational constant by G, and the solar mass by $M_\odot$.
Indices. Greek letters denote spacetime indices, while Latin letters denote space indices.
Minkowski metric tensor.

$$\eta_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$$

Four-vector.

$$x^\mu = (ct, \boldsymbol{x}) = (ct, x, y, z)$$
$$\mathrm{d}^4x = c\mathrm{d}t\mathrm{d}^3x$$

Partial derivative.

$$f_{,\mu} = \partial_\mu f = \frac{\partial f}{\partial x^\mu}$$

Time derivative.

$$\dot{f} = \frac{\partial f}{\partial t}$$

D'Alambert operator.

$$\Box = \eta^{\mu\nu}\partial_\mu\partial_\nu$$

Christoffel symbol.

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

Covariant derivative. We denote the covariant derivative with a semicolon. For instance, the covariant derivative for a first-order contravariant vector is expressed as

$$A^\mu{}_{;\nu} = A^\mu{}_{,\nu} + A^\alpha\Gamma^\mu{}_{\alpha\nu}$$

Riemann tensor.

$$R^\mu{}_{\nu\rho\sigma} = \Gamma^\mu{}_{\nu\sigma,\rho} - \Gamma^\mu{}_{\nu\rho,\sigma} + \Gamma^\mu{}_{\alpha\rho}\Gamma^\alpha{}_{\nu\sigma} - \Gamma^\mu{}_{\alpha\sigma}\Gamma^\alpha{}_{\nu\rho}$$

Ricci tensor.

$$R_{\mu\nu} = R^\alpha{}_{\mu\alpha\nu}$$

Ricci scalar.

$$R = g^{\mu\nu}R_{\mu\nu}$$

Fourier transform. Our conventions on the n-dimensional Fourier transform are

$$F(x) = \frac{1}{(2\pi)^n} \int \mathrm{d}^n k \; \tilde{F}(k) e^{ikx}$$
$$\tilde{F}(k) = \int \mathrm{d}^n x \; F(x) e^{-ikx}$$

Particulary, for a function of time,

$$F(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega \; \tilde{F}(\omega) e^{-i\omega t}$$
$$\tilde{F}(\omega) = \int_{-\infty}^{\infty} \mathrm{d}t \; F(t) e^{i\omega t}$$

Acronyms & Abbreviations

e.g.	for example (exempli gratia)
i.e.	that is (id est)
i.i.d.	independent and identically distributed
BBH	binary black hole
BNS	binary neutron star
CAM	class activation mapping
CBC	compact binary coalescence
CCSN	core-collapse supernova
CNN	convolutional neural network
GW	gravitational wave
H1	LIGO Hanford detector
L1	LIGO Livingston detector
LIGO	Laser Interferometer Gravitational-wave Observatory
NSBH	neutron star-black hole
O1	the first observing run
O2	the second observing run
O3	the third observing run
O4	the fourth observing run
ReLU	rectified linear unit
ResNet	residual neural network
RMSE	root mean squared error
ROC	receiver operating characteristic
SASI	standing accretion shock instability
SNR	signal-to-noise ratio
STFT	short-time Fourier transform
t-SNE	t-distributed stochastic neighbor embedding
V1	Virgo detector

Chapter 1

Introduction

Throughout history, humans have learned a great deal by observing the Universe. From observing stars with the naked eye to the groundbreaking development of telescopes in the late 16th century, our understanding of the Universe has improved significantly. Technological advancements have led to the development of various instruments, including radio telescopes, X-ray telescopes, and gamma-ray telescopes, which have revealed important discoveries such as the cosmic microwave background. The recently launched James Webb Space Telescope [4] enables us to explore the distant history of the Universe at near-infrared wavelengths. X-ray and gamma-ray observations have revealed high-energy events such as supernovae and pulsars. Although electromagnetic observations have their strengths, there are limitations, particularly in the study of the early Universe, due to potential challenges of photon propagation in high-energy conditions. A major milestone occurred on September 14, 2015, when the Advanced Laser Interferometer Gravitational-wave Observatory (Advanced LIGO) [5] detected the first gravitational wave (GW), introducing a new approach to observing the Universe [6].

GWs are ripples in spacetime caused by accelerating objects with asymmetrical mass distribution, initially theorized to exist by Einstein as a consequence of the general theory of relativity [7, 8]. Due to their minuscule nature, only GWs generated by astrophysical phenomena are detectable on Earth, posing a significant challenge to their detection. In the 1960s, Weber carried out pioneering experiments using resonant bar detectors made of aluminum [9]. He claimed to have detected a GW by observing a coincident signal in two detectors [10], but subsequent follow-up experiments and analyses considered the event to be a false positive [11]. Subsequently, in the 1970s, Hulse and Taylor found a binary pulsar system known as PSR 1913+16 [12], and by observing its orbit, they provided indirect evidence for the emission of GWs from this system [13]. Their research not only confirmed the existence of GWs, but also provided substantial support for the theory of general relativity. In the 1980s, the concept of using laser interferometry to detect GWs over a broad frequency range began to gain momentum. By the 2000s, first-generation detectors including TAMA300 in Tokyo, Japan [14], GEO600 near Hanover, Germany [15], Virgo near Pisa, Italy [16], and two LIGO detectors in Hanford, Washington and Livingston, Louisiana in the United States [17], were constructed. However, these detectors did not succeed in detecting GW events.

Significant advancements occurred in the 2010s with the introduction of upgraded second-generation detectors. Specifically, the two LIGO detectors were upgraded to Advanced LIGO, while Virgo underwent an upgrade to Advanced Virgo [18]. In addition, a

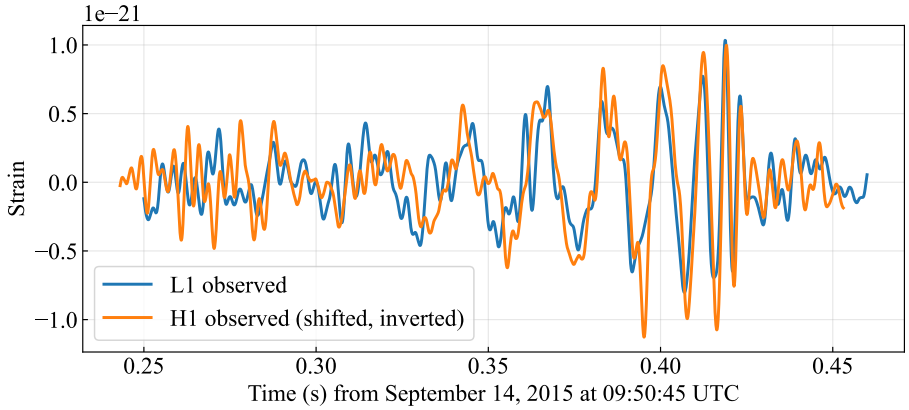


FIG. 1.1. The GW event GW150914 observed by the LIGO Hanford (H1) and Livingston (L1) detectors. Each strain time series is filtered with a 35-350 Hz band-pass filter. The strain at the H1 detector is shifted in time by 6.9 ms and inverted. Data taken from Ref. [36].

new detector named KAGRA was constructed in Gifu, Japan [19]. The first direct GW detection was achieved by the two detectors of Advanced LIGO, shown in Fig. 1.1. The detected GW, named GW150914, was generated from a binary black hole (BBH) merger with component masses of $36M_\odot$ and $29M_\odot$ [6]. In 2017, the two LIGO detectors and Advanced Virgo detected a GW signal from a binary neutron star (BNS) merger with a total mass of $2.7M_\odot$, which is referred to as GW170817 [20]. The source was localized within a sky region of 28 deg^2 based on data from the three detectors. This precise localization enabled a multi-messenger observation of the event [21]. Shortly after the event, approximately 1.7 s following the merger, the FERMI and INTEGRAL telescopes detected a gamma-ray burst, known as GRB170817A [22, 23]. The optical and infrared counterpart associated with this event was identified in the direction of the Hydra constellation, specifically in the galaxy NGC4993. These GW and electromagnetic observations of the BNS merger provided constraints on the tidal deformability and the nuclear equation of state of neutron stars [24, 25], ruled out some gravity theories [26, 27], and provided new insights into various aspects, such as the Hubble constant [28] and kilonovae [29]. In the third observing run (O3), two neutron star-black hole (NSBH) mergers were detected, completing the expected set of detectable compact binary systems [30]. In the course of three observing runs, Advanced LIGO and Advanced Virgo reported 90 confident GW detections from compact binary coalescences (CBCs) in total [31–34]. In the fourth observing run (O4), which commenced in May 2023, more than 80 CBC events have already been detected [35].

As detector sensitivity improves and new detectors are constructed, the frequency of GW detection is expected to increase. Consequently, there is a growing need for computationally efficient algorithms to detect and analyze these signals. The current search method for CBC signals is called *matched filtering* [37], an optimal method for detecting a known signal in stationary Gaussian noise. It works by taking the correlation of theoretical waveforms, drawn from a set of candidate signals, with the detector strain. However, it introduces a significant computational cost as the number of waveform parameters increases. Reducing computational costs for the detection and source localization are crucial

for enabling follow-up observations by electromagnetic telescopes.

Another expected GW source is the core-collapse supernova (CCSN). This phenomenon is an explosion of massive stars at the end of their lifetimes, leading to the formation of neutron stars or black holes. While supernova explosions have been frequently observed through electromagnetic waves, we cannot observe electromagnetic waves produced near the central engine due to the large scattering cross section. Since GWs and neutrinos can propagate to us without losing physical information from the central engine, detecting them is essential for understanding the explosion mechanism and revealing the internal structure of neutron stars. Neutrinos from a supernova were observed in 1987 from a supernova event SN1987A in the Large Magellanic Cloud at 51 kpc away from Earth [38, 39]. Observations of this event by neutrinos and electromagnetic waves confirmed the theory of stellar evolution process, and revealed that about 99% of the gravitational binding energy ($\sim 10^{53}$ erg) is released as neutrinos. However, the details of the explosion mechanism are not yet fully understood, and GW signals from CCSNe have yet to be detected. In recent years, multi-dimensional numerical simulations of supernova explosions have been intensively studied. Their GW signals have stochastic nature due to the hydrodynamical instabilities [40], making it difficult to accurately model their waveforms. Consequently, the matched-filtering technique is not practical in the search for GWs from CCSNe, and alternative detection methods are required.

Machine learning has emerged as a new approach in GW data analysis, and has been actively studied over the past several years [41, 42]. In particular, deep learning, including convolutional neural networks (CNNs) has gained momentum, following the success in fields such as image recognition and natural language processing [43]. A key characteristic of deep learning models is their ability to make predictions in a very short time, which may be valuable for low-latency searches. The ability to learn robust features from data sets could be beneficial in searching for GW signals from unmodeled sources. The applications in GW physics range from CBC search [44–46], CCSN search [46, 47], parameter estimation [48–50], classification of transient noises [51–53], and control of interferometers [54], to name a few. While deep learning has demonstrated robust performance on various tasks, the complex nature of its models, characterized by an extremely large number of parameters, poses a challenge in understanding their decision-making processes. In response to this challenge, the field of explainable artificial intelligence [55] has gained prominence, aiming to make models' decisions transparent and interpretable. Specifically, in the context of CNNs, efforts have been made to develop techniques that attempt to understand the decision-making process by mapping the output of the network back into the input space to identify the specific input components that were discriminative in producing the output. In our research, we utilize this techniques to GW detection and classification models.

Another data analysis technique used in our study is the Hilbert-Huang transform [56]. While time-frequency maps are typically generated using methods such as short-time Fourier transform or wavelet transforms, their limitations arise from the uncertainty relationship between duration and bandwidth. In this context, the Hilbert-Huang transform holds a distinct advantage as it is not affected by this relationship, given its ability to compute instantaneous frequency for each data point. This feature makes it a valuable

tool for delving into the underlying physics of GW signals. Previous research [57] utilized this technique for determining frequencies of GW modes from hydrodynamic instabilities in CCSNe. In this study, we perform a basic study towards extracting GW modes arising from oscillations of proto-neutron stars to estimate their parameters.

Part I of the thesis will focus on the theory of GW astronomy and GW data analysis method used in the research projects. In Chapter 2, we provide the principles of GW physics. Chapter 3 describes the theory of deep learning and their applications in GW data analysis. The fundamentals of the Hilbert-Huang transform and its applications in GW astronomy will be discussed in Chapter 4.

Our research results are presented in Part II. In Chapter 5, we describe our study published in Ref. [2]. In the study, we apply explainable artificial intelligence techniques to a CNN classifier of GW signals from CCSNe. Chapter 6 is a reproduction of the paper [3]. This study compares different CNN detection models of GWs from CBC sources through feature attribution analysis. In Chapter 7, we present our results on the applications of the Hilbert-Huang transform of GW signals to parameter estimation of proto-neutron stars.

Part I

Theoretical Foundation

Chapter 2

Gravitational Wave

This chapter covers the principles of GW theory, mainly following the textbook by Maggiore [58]. We derive the basic characteristics of GWs and describe their detectors and the methods used to analyze their data, as an introduction to the research presented in the subsequent chapters.

2.1 Derivation from general relativity

2.1.1 Propagation

The gravitational action is given by $S = S_{\rm EH} + S_{\rm M}$, where $S_{\rm M}$ denotes the matter action, and $S_{\rm EH}$ represents the Einstein-Hilbert action, defined by

$$S_{\rm EH} = \frac{c^3}{16\pi G}\int \mathrm{d}^4x\ \sqrt{-g}R. \tag{2.1}$$

Let us consider the vacuum spacetime, where $S_{\rm M} = 0$. We express the metric tensor as the sum of the Minkowski metric $\eta_{\mu\nu} = \mathrm{diag}(-1,1,1,1)$ and a perturbative term as

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{2.2}$$

and expand the action to quadratic order in $h_{\mu\nu}$. The Ricci scalar R is expanded as

$$\begin{aligned} R &= g^{\mu\nu}R_{\mu\nu} \\ &= \eta^{\mu\nu}R^{(1)}_{\mu\nu} - h^{\mu\nu}R^{(1)}_{\mu\nu} + \eta^{\mu\nu}R^{(2)}_{\mu\nu} + \mathcal{O}(h^3) \end{aligned} \tag{2.3}$$

where $R^{(1)}_{\mu\nu}$ and $R^{(2)}_{\mu\nu}$ are the linear and quadratic terms of the Ricci tensor for h, respectively. These terms are given by (refer to Appendix A)

$$R^{(1)}_{\mu\nu} = \frac{1}{2}(-h_{,\mu\nu} - h_{\mu\nu,\alpha}{}^{\alpha} + h_{\alpha\mu,\nu}{}^{\alpha} + h_{\alpha\nu,\mu}{}^{\alpha}), \tag{2.4}$$

$$\begin{aligned} R^{(2)}_{\mu\nu} = \frac{1}{2}\Bigg[&\frac{1}{2}h_{\alpha\beta,\mu}h^{\alpha\beta}{}_{,\nu} + h^{\alpha\beta}(h_{\alpha\beta,\mu\nu} + h_{\mu\nu,\alpha\beta} - h_{\alpha\mu,\nu\beta} - h_{\alpha\nu,\mu\beta}) \\ &+h_{\nu}{}^{\alpha,\beta}(h_{\alpha\mu,\beta} - h_{\beta\mu,\alpha}) - \left(h^{\alpha\beta}{}_{,\beta} - \frac{1}{2}h^{,\alpha}\right)(h_{\alpha\mu,\nu} + h_{\alpha\nu,\mu} - h_{\mu\nu,\alpha})\Bigg]. \end{aligned} \tag{2.5}$$

By using the expansion of $\sqrt{-g}$:

$$\sqrt{-g} = 1 + \frac{1}{2}h + \mathcal{O}(h^2), \tag{2.6}$$

we obtain the equation

$$\sqrt{-g}R = \eta^{\mu\nu}R^{(1)}_{\mu\nu} + \frac{h}{2}\eta^{\mu\nu}R^{(1)}_{\mu\nu} - h^{\mu\nu}R^{(1)}_{\mu\nu} + \eta^{\mu\nu}R^{(2)}_{\mu\nu} + \mathcal{O}(h^3). \tag{2.7}$$

Substituting Eqs. (2.4), (2.5), and (2.7) into Eq. (2.1) yields the expansion of the Einstein-Hilbert action to quadratic order in h, expressed as

$$S_{\mathrm{EH}} = -\frac{c^3}{16\pi G}\int \mathrm{d}^4x\Big(h_{\alpha\beta,\mu}h^{\alpha\beta,\mu} - h_{,\mu}h^{,\mu} + 2h^{\mu\nu}{}_{,\mu}h_{,\nu} - 2h^{\mu\nu}{}_{,\mu}h^{\rho}{}_{\nu,\rho} + \mathcal{O}(h^3)\Big). \tag{2.8}$$

The corresponding Lagrangian density is

$$\mathcal{L} = -\frac{c^4}{16\pi G}\Big(h_{\alpha\beta,\mu}h^{\alpha\beta,\mu} - h_{,\mu}h^{,\mu} + 2h^{\mu\nu}{}_{,\mu}h_{,\nu} - 2h^{\mu\nu}{}_{,\mu}h^{\rho}{}_{\nu,\rho} + \mathcal{O}(h^3)\Big). \tag{2.9}$$

Here, the linear term vanishes when expanding around a classical solution. Variating this action with respect to h leads to the linearized Einstein equation in the vacuum,

$$-h_{,\mu\nu} + h_{\mu\rho,\nu}{}^{\rho} + h_{\nu\rho,\mu}{}^{\rho} - \Box h_{\mu\nu} - \eta_{\mu\nu}(h_{\rho\sigma}{}^{,\rho\sigma} - \Box h) = 0. \tag{2.10}$$

Introducing

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h \tag{2.11}$$

and rewriting Eq. (2.10) with $\bar{h}_{\mu\nu}$ results in a simpler equation,

$$\Box\bar{h}_{\mu\nu} + \eta_{\mu\nu}\bar{h}_{\rho\sigma}{}^{,\rho\sigma} - \bar{h}_{\mu\rho,\nu}{}^{\rho} - \bar{h}_{\nu\rho,\mu}{}^{\rho} = 0. \tag{2.12}$$

While choosing a frame in which the linearization Eq. (2.2) holds, there still remains a residual gauge symmetry. Consider a transformation,

$$x^{\mu} \to x'^{\mu} = x^{\mu} + \xi^{\mu}(x). \tag{2.13}$$

Under this transformation, $h_{\mu\nu}$ and $\bar{h}_{\mu\nu}$ are transformed as

$$h_{\mu\nu}(x) \to h'_{\mu\nu}(x') = h_{\mu\nu}(x) - (\xi_{\nu,\mu} + \xi_{\mu,\nu}), \tag{2.14}$$

$$\bar{h}_{\mu\nu}(x) \to \bar{h}'_{\mu\nu}(x') = \bar{h}_{\mu\nu}(x) - (\xi_{\nu,\mu} + \xi_{\mu,\nu} - \eta_{\mu\nu}\xi^{\rho}{}_{,\rho}). \tag{2.15}$$

We choose the Lorentz gauge, defined as

$$\bar{h}_{\mu\nu}{}^{,\nu} = 0. \tag{2.16}$$

We can impose this condition by choosing ξ that satisfies

$$\bar{h}_{\mu\nu}{}^{,\nu} \to (\bar{h}_{\mu\nu}{}^{,\nu})' = \bar{h}_{\mu\nu}{}^{,\nu} - \Box\xi_{\mu} = 0. \tag{2.17}$$

Imposing the Lorentz gauge simplifies the linearized Einstein equation to a simple wave equation,

$$\Box \bar{h}_{\mu\nu} = 0. \tag{2.18}$$

Solutions of Eq. (2.18) are plane waves, represented by

$$\bar{h}_{\mu\nu} = A_{\mu\nu} \exp(ik_\rho x^\rho), \tag{2.19}$$

where $k_\rho k^\rho = 0$. Since $h_{\mu\nu}$ is symmetric, $A_{\mu\nu}$ is also symmetric, exhibiting ten independent components. Enforcing the Lorentz gauge Eq. (2.16) reduces the number of independent components to six. Additionally, ξ_μ comprises four independent components, leaving two independent components in $A_{\mu\nu}$. ξ_μ also satisfies the wave equation and can be expressed as

$$\xi_\mu = B_\mu \exp(ik_\rho x^\rho). \tag{2.20}$$

From Eq. (2.14), the transformation of $A_{\mu\nu}$ is given by

$$A_{\mu\nu} \to A'_{\mu\nu} = A_{\mu\nu} - iB_\mu k_\nu - iB_\nu k_\mu + i\eta_{\mu\nu} B_\rho k^\rho. \tag{2.21}$$

Let us consider a wave traveling in the z-direction,

$$k^\mu = (k, 0, 0, k). \tag{2.22}$$

By choosing B_μ such that

$$B_0 = \frac{i(2A_{00} + A_{11} + A_{22})}{4k}, \tag{2.23}$$

$$B_1 = \frac{iA_{01}}{k}, \tag{2.24}$$

$$B_2 = \frac{iA_{02}}{k}, \tag{2.25}$$

$$B_3 = \frac{i(2A_{00} - A_{11} - A_{22})}{4k}, \tag{2.26}$$

two conditions

$$A'_{0\nu} = 0, \tag{2.27}$$

$$\eta^{\mu\nu} A'_{\mu\nu} = 0, \tag{2.28}$$

are satisfied, in addition to the Lorentz-gauge condition. These gauge conditions are termed as the *transverse traceless gauge*. With the transverse traceless gauge, since the trace of the transformed tensor is zero, $\bar{h}'_{\mu\nu} = h'_{\mu\nu}$ holds, denoted as $h^{\mathrm{TT}}_{\mu\nu}$. $h^{\mathrm{TT}}_{\mu\nu}$ are expressed by two independent components h_+ and $h_\times$ as

$$h^{\mathrm{TT}}_{ij} = (h_+ e^+_{ij} + h_\times e^\times_{ij}) \exp(ik_\rho x^\rho), \tag{2.29}$$

where

$$e_{ij}^{+} = \hat{e}_1 \otimes \hat{e}_1 - \hat{e}_2 \otimes \hat{e}_2, \tag{2.30}$$

$$e_{ij}^{\times} = \hat{e}_1 \otimes \hat{e}_2 + \hat{e}_2 \otimes \hat{e}_1. \tag{2.31}$$

h_+ and $h_\times$ are referred to as the *plus mode* and *cross mode* of GWs, respectively. In Fig. 2.1, we illustrate the effect of these two modes when GWs are normally incident to the plane.

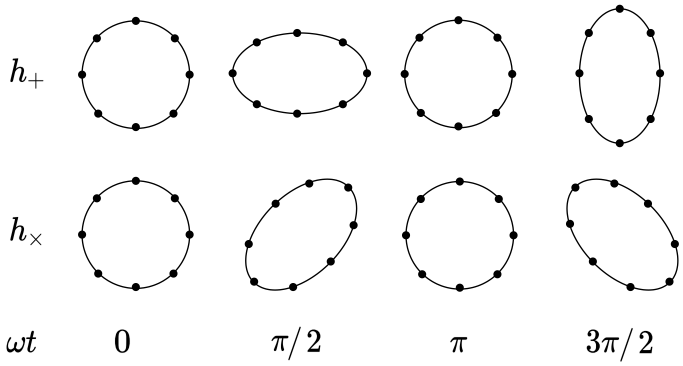


FIG. 2.1. Illustration of the effect of the plus and cross modes of GWs on a circular ring of free masses.

2.1.2 Generation

This section delves into the generation of GWs, demonstrating that the amplitude of GWs can be expressed through the second time derivative of the quadrupole moment. We initiate the discussion by considering the linearized Einstein equation,

$$\Box \bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4} T_{\mu\nu}. \tag{2.32}$$

By employing the Green's function, the solution to this wave equation can be expressed as

$$\bar{h}_{\mu\nu}(t, \boldsymbol{x}) = \frac{4G}{c^4} \int \mathrm{d}^3 x' \frac{1}{|\boldsymbol{x} - \boldsymbol{x}'|} T_{\mu\nu}\left(t - \frac{|\boldsymbol{x} - \boldsymbol{x}'|}{c}, \boldsymbol{x}'\right). \tag{2.33}$$

In the transverse traceless gauge, h_{ij} can be expressed as

$$h_{ij}^{\mathrm{TT}}(t, \boldsymbol{x}) = \frac{4G}{c^4} \Lambda_{ij,kl} \int \mathrm{d}^3 x' \frac{1}{|\boldsymbol{x} - \boldsymbol{x}'|} T_{kl}\left(t - \frac{|\boldsymbol{x} - \boldsymbol{x}'|}{c}, \boldsymbol{x}'\right) \tag{2.34}$$

where $\Lambda_{ij,kl}$ is the projecting operator into the transverse traceless gauge, given by

$$\Lambda_{ij,kl} = P_{ik} P_{jl} - \frac{1}{2} P_{ij} P_{kl}, \tag{2.35}$$

where

$$P_{ij} = \delta_{ij} - n_i n_j. \tag{2.36}$$

Let us consider the multipole expansion of the right-hand side of the equation. The velocity of the source can be expressed as $v \sim \omega_s d$, where the typical angular frequency of

the source is ω_s and the magnitude is d. Consequently, the frequency ω of the radiating GW is of the order of ω_s, leading to $\omega \sim v/d$, or

$$\lambda \sim \frac{c}{v} d. \tag{2.37}$$

Hence, in a non-relativistic system where $v \ll c$ holds, the wavelength of GWs is significantly larger than that of the source. In such cases, understanding the details of the motion inside the source is not necessary, allowing consideration of only the lowest-order terms of the multipole moment. If $r = |\boldsymbol{x}|$ is significantly larger than the size of the source d, then we can use the approximation,

$$\left|\boldsymbol{x} - \boldsymbol{x}'\right| \simeq r - \boldsymbol{x}' \cdot \hat{\boldsymbol{x}}. \tag{2.38}$$

Thus, Eq. (2.34) can be approximated as

$$h_{ij}^{\mathrm{TT}}(t, \boldsymbol{x}) = \frac{1}{r} \frac{4G}{c^4} \Lambda_{ij,kl} \int \mathrm{d}^3 x' \, T_{kl}\left(t - \frac{r}{c} + \frac{\boldsymbol{x}' \cdot \hat{\boldsymbol{x}}}{c}, \boldsymbol{x}'\right). \tag{2.39}$$

Using the Taylor expansion of the integrand,

$$\begin{aligned} T_{kl}\left(t - \frac{r}{c} + \frac{\boldsymbol{x}' \cdot \hat{\boldsymbol{x}}}{c}, \boldsymbol{x}'\right) = T_{kl}\left(t - \frac{r}{c}, \boldsymbol{x}'\right) + \boldsymbol{x}' \cdot \hat{\boldsymbol{x}} \, \partial_0 T_{kl}\left(t - \frac{r}{c}, \boldsymbol{x}'\right) \\ + \frac{1}{2} (\boldsymbol{x}' \cdot \hat{\boldsymbol{x}})^2 \partial_0^2 T_{kl}\left(t - \frac{r}{c}, \boldsymbol{x}'\right) + \cdots, \end{aligned} \tag{2.40}$$

Eq. (2.39) can be expressed as

$$h_{ij}^{\mathrm{TT}}(t, \boldsymbol{x}) = \frac{1}{r} \frac{4G}{c^4} \Lambda_{ij,kl} \left[S^{kl} + \hat{x}_m \partial_0 S^{kl,m} + \frac{1}{2} \hat{x}_m \hat{x}_n \partial_0^2 S^{kl,mn} \right]_{\mathrm{ret}}, \tag{2.41}$$

where

$$S^{ij}(t) = \int \mathrm{d}^3 x \, T^{ij}(t, \boldsymbol{x}), \tag{2.42}$$

$$S^{ij,k}(t) = \int \mathrm{d}^3 x \, T^{ij}(t, \boldsymbol{x}) x^k, \tag{2.43}$$

$$S^{ij,kl}(t) = \int \mathrm{d}^3 x \, T^{ij}(t, \boldsymbol{x}) x^k x^l. \tag{2.44}$$

are the momenta of T^{ij}, and 'ret' means that the momenta and their derivatives are evaluated at $(t - r/c, \boldsymbol{x}')$. Furthermore, we define the momentum related to the energy

density T^{00}/c^2 and momentum density T^{0i}/c,

$$M = \frac{1}{c^2}\int \mathrm{d}^3x\, T^{00}(t,\boldsymbol{x}), \tag{2.45}$$

$$M^i = \frac{1}{c^2}\int \mathrm{d}^3x\, T^{00}(t,\boldsymbol{x})x^i, \tag{2.46}$$

$$M^{ij} = \frac{1}{c^2}\int \mathrm{d}^3x\, T^{00}(t,\boldsymbol{x})x^ix^j, \tag{2.47}$$

$$M^{ijk} = \frac{1}{c^2}\int \mathrm{d}^3x\, T^{00}(t,\boldsymbol{x})x^ix^jx^k, \tag{2.48}$$

$$P^i = \frac{1}{c}\int \mathrm{d}^3x\, T^{0i}(t,\boldsymbol{x}), \tag{2.49}$$

$$P^{i,j} = \frac{1}{c}\int \mathrm{d}^3x\, T^{0i}(t,\boldsymbol{x})x^j, \tag{2.50}$$

$$P^{i,jk} = \frac{1}{c}\int \mathrm{d}^3x\, T^{0i}(t,\boldsymbol{x})x^jx^k. \tag{2.51}$$

Since we are considering a linear approximation of the gravitational field, the energy-momentum tensor satisfies the conservation law,

$$T^{\mu\nu}{}_{,\nu} = 0. \tag{2.52}$$

Thus, within a sufficiently large volume V away from the source, we have

$$c\dot{M} = \int \mathrm{d}^3x\, \partial_0 T^{00} = -\int \mathrm{d}^3x\, \partial_i T^{0i} = -\int_{\partial V} \mathrm{d}S_i\, T^{0i} = 0. \tag{2.53}$$

The third equality is based on Gauss's theorem. A similar calculation yields the following relationships:

$$\dot{M} = 0 \tag{2.54}$$

$$\dot{M}^i = P^i \tag{2.55}$$

$$\dot{M}^{ij} = P^{i,j} + P^{j,i} \tag{2.56}$$

$$\dot{M}^{ijk} = P^{i,jk} + P^{j,ki} + P^{k,ij} \tag{2.57}$$

$$\dot{P}^i = 0 \tag{2.58}$$

$$\dot{P}^{i,j} = S^{ij} \tag{2.59}$$

$$\dot{P}^{i,jk} = S^{ij,k} + S^{ik,j} \tag{2.60}$$

By differentiating Eq. (2.56) with respect to time and using the symmetry of S^{ij} along with Eq. (2.59), we obtain

$$\ddot{M}^{ij} = 2S^{ij}. \tag{2.61}$$

Using this equation, the primary term in Eq. (2.41) is described as

$$h_{ij}^{\mathrm{TT}}(t,\boldsymbol{x}) = \frac{1}{r}\frac{2G}{c^4}\Lambda_{ij,kl}\ddot{M}^{kl}(t-r/c). \tag{2.62}$$

By introducing the traceless quadrupole moment,

$$Q^{kl} = M^{kl} - \frac{1}{3}\delta^{kl} M_{ii}, \tag{2.63}$$

h_{ij}^{TT} can be rewritten as

$$h_{ij}^{\mathrm{TT}}(t, \boldsymbol{x}) = \frac{1}{r}\frac{2G}{c^4}\Lambda_{ij,kl}\ddot{Q}_{kl}(t - r/c) \tag{2.64}$$

$$= \frac{1}{r}\frac{2G}{c^4}\ddot{Q}_{ij}^{\mathrm{TT}}(t - r/c). \tag{2.65}$$

This expression is known as the quadrupole formula of GWs.

2.1.3 Energy-momentum tensor

The energy-momentum tensor carried by GWs is described by

$$t^{\mu\nu} = \left\langle -\frac{\partial \mathcal{L}}{\partial(h_{\alpha\beta,\mu})} h_{\alpha\beta}{}^{,\nu} + \eta^{\mu\nu}\mathcal{L} \right\rangle, \tag{2.66}$$

where $\langle\cdot\rangle$ is a spatial average over several reduced wavelength, and $\mathcal{L}$ is the Lagrangian density defined in Eq. (2.9). In the transverse traceless gauge, by inserting the Lagrangian density and using the equation of motion $\Box h_{\mu\nu} = 0$ and the gauge conditions, we obtain the gauge free expression

$$t^{ij} = \frac{c^4}{32\pi G}\langle h^{kl,i} h_{kl}{}^{,j}\rangle. \tag{2.67}$$

The gauge-invariant energy density is

$$t^{00} = \frac{c^2}{32\pi G}\langle \dot{h}_{ij}\dot{h}_{ij}\rangle$$

$$= \frac{c^2}{16\pi G}\langle \dot{h}_+^2 + \dot{h}_\times^2\rangle. \tag{2.68}$$

The propagating GW carries away an energy flux

$$\frac{\mathrm{d}E}{\mathrm{d}t} = c\int \mathrm{d}A\ t^{00}$$

$$= \frac{c^3 r^2}{16\pi G}\int \mathrm{d}\Omega\ \langle \dot{h}_+^2 + \dot{h}_\times^2\rangle. \tag{2.69}$$

By using the quadrupole formula, we have

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{G}{8\pi c^5}\int \mathrm{d}\Omega\ \Lambda_{ij,kl}\langle \dddot{Q}_{ij}\dddot{Q}_{kl}\rangle, \tag{2.70}$$

and using the integral of the Lambda operator:

$$\int \mathrm{d}\Omega\ \Lambda_{ij,kl} = \frac{2\pi}{5}(11\delta_{ik}\delta_{jl} - 4\delta_{ij}\delta_{kl} + \delta_{il}\delta_{jk}), \tag{2.71}$$

we obtain the expression of the energy flux,

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{G}{5c^5}\langle \dddot{Q}_{ij}\, \dddot{Q}_{ij}\rangle. \tag{2.72}$$

2.2 Detector

This chapter describes the basics of the Michelson interferometer and its interaction with GWs, along with the summary of current and future GW detectors.

2.2.1 Michelson interferometer

A schematic diagram of the Michelson interferometer is shown in Fig. 2.2. A laser beam is split into two directions at the beam splitter, travels along the x and y arms, reflects off a mirror, returns to the beam splitter, and interferes. GWs change the fringes of the Michelson interferometer by differentially changing the length of the space at 90 degrees to each other. Therefore, GWs can be detected from changes in the intensity of the interferogram of the Michelson interferometer.

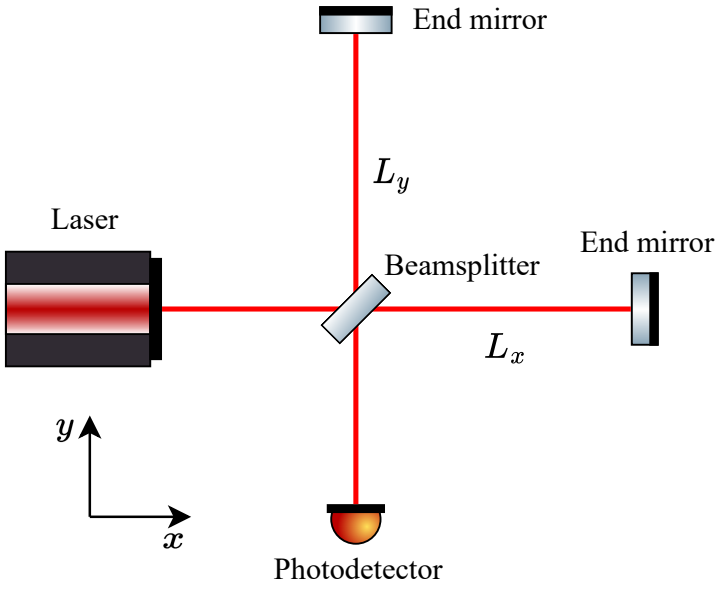


FIG. 2.2. Schematic diagram of a Michelson interferomter.

Let L_x and L_y be the lengths of the arms of the Michelson interferometer in the x and y directions, respectively, and consider the response of the Michelson interferometer to the plus mode of a GW in the transverse traceless gauge coming from the z direction. The spacetime interval is given by

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + (1 + h_+(t))\mathrm{d}x^2 + (1 - h_+(t))\mathrm{d}y^2 + \mathrm{d}z^2. \tag{2.73}$$

Photons travel along null geodesics, i.e. $\mathrm{d}s^2 = 0$. For the light in the x arm, to first order in h_+, we have

$$\mathrm{d}x = \pm c\,\mathrm{d}t\left(1 - \frac{1}{2}h_+(t)\right), \tag{2.74}$$

where the plus sign represents light traveling from the beam splitter to the mirror, and the minus sign represents the opposite direction. Consider a photon leaving the beam splitter at a time t_0 and reaching the mirror in the x axis at a time t_1. Integrating Eq. (2.74) yields

$$L_x = c(t_1 - t_0) - \frac{c}{2}\int_{t_0}^{t_1} \mathrm{d}t'\ h_+(t'). \tag{2.75}$$

Suppose it returns to the beam splitter at a time t,

$$-L_x = -c(t - t_1) + \frac{c}{2}\int_{t_1}^{t} \mathrm{d}t'\ h_+(t'). \tag{2.76}$$

Adding the equations above, and considering t_0 in the integral interval to the first order in h_+ as $t - 2L_x/c$, the time Δt_x for a round trip of the x arm is given by

$$\begin{aligned}\Delta t_x &= t - t_0 \\ &= \frac{2L_x}{c} + \frac{1}{2}\int_{t_0}^{t} \mathrm{d}t'\ h_+(t') \\ &= \frac{2L_x}{c} + \frac{1}{2}\int_{t-2L_x/c}^{t} \mathrm{d}t'\ h_+(t').\end{aligned} \tag{2.77}$$

Similarly, in the y direction, we have

$$\Delta t_y = \frac{2L_y}{c} - \frac{1}{2}\int_{t-2L_y/c}^{t} \mathrm{d}t'\ h_+(t') \tag{2.78}$$

The phase change $\Delta\phi$ is then expressed as

$$\Delta\phi = \Omega_0(\Delta t_x - \Delta t_y) \tag{2.79}$$

$$= \frac{2(L_x - L_y)\Omega_0}{c} + \Omega_0\int_{t-2L/c}^{t} \mathrm{d}t'\ h_+(t') \tag{2.80}$$

where Ω_0 is the angular frequency of the laser, and we assumed $L_x \simeq L_y \simeq L$ in the second term. The second term is the phase difference induced by GWs:

$$\Delta\phi_{\mathrm{GW}} = \Omega_0\int_{t-2L/c}^{t} \mathrm{d}t'\ h_+(t'). \tag{2.81}$$

Let $\tilde{h}(\omega)$ be the Fourier transform of $h_+(t)$. Then the phase difference is

$$\begin{aligned}\Delta\phi_{\mathrm{GW}} &= \frac{\Omega_0}{2\pi}\int_{t-2L/c}^{t} \mathrm{d}t' \int_{-\infty}^{\infty} \mathrm{d}\omega\ \tilde{h}(\omega)e^{-i\omega t'} \\ &= -\int_{-\infty}^{\infty} \mathrm{d}\omega\ \frac{\Omega_0}{\pi\omega}\sin\left(\frac{\omega L}{c}\right)e^{i\omega L/c}\tilde{h}(\omega)e^{-i\omega t}.\end{aligned} \tag{2.82}$$

The amplitude of the frequency response function takes the maximum at $\omega L/c = \pi/2$. From this equation, the optimal arm length for detecting GWs of 100 Hz is calculated as 750 km, but it is practically impossible to install such a large interferometer on the ground. Therefore, in the current detectors, optical resonators are applied in the interferomters to

extend the effective arm length.

2.2.2 Antenna pattern

Here, let us consider GWs from arbitrary directions. The response of a detector against a GW h^{ij} is generally expressed using a detector tensor D_{ij} as [58]

$$h = D_{ij}h^{ij}. \tag{2.83}$$

The detector tensor of an interferometer that has arms in x and y axes is

$$D_{ij} = \frac{1}{2}(\hat{\boldsymbol{x}}_i\hat{\boldsymbol{x}}_j - \hat{\boldsymbol{y}}_i\hat{\boldsymbol{y}}_j) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}_{ij}. \tag{2.84}$$

Hence, the response in this case is expressed as $h = (h_{11} - h_{22})/2$. The phase difference of this GW can be obtained by replacing h_+ with h in Eq. (2.81). However, h_{11} and h_{22} should be expressed in the detector frame, not in the source frame. We transform a GW in the source frame (x', y', z') into the detector frame (x, y, z). In the source frame, a GW signal in the transverse traceless gauge is written as

$$h'_{ij} = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}_{ij}. \tag{2.85}$$

We define three angles, θ, ϕ, and ψ between the two frames as illustrated in Fig. 2.3. The rotation matrix from (x', y', z') to (x, y, z) is

$$\begin{aligned} R &= R_z^{-1}(\phi)R_y^{-1}(\theta)R_z^{-1}(\psi) \\ &= \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \cos\theta\cos\phi\cos\psi - \sin\phi\sin\psi & \cos\theta\cos\phi\sin\psi + \sin\phi\cos\psi & \sin\theta\cos\phi \\ -\cos\theta\sin\phi\cos\psi - \cos\phi\sin\psi & -\cos\theta\sin\phi\sin\psi + \cos\phi\cos\psi & -\sin\theta\sin\phi \\ -\sin\theta\cos\psi & -\sin\theta\sin\psi & \cos\theta \end{pmatrix}, \end{aligned} \tag{2.86}$$

and a GW in the source frame h'_{kl} is transformed into the detector frame h_{ij} by

$$h_{ij} = \mathcal{R}_{ik}\mathcal{R}_{jl}h'_{kl}. \tag{2.87}$$

By the straightforward matrix calculations, we obtain

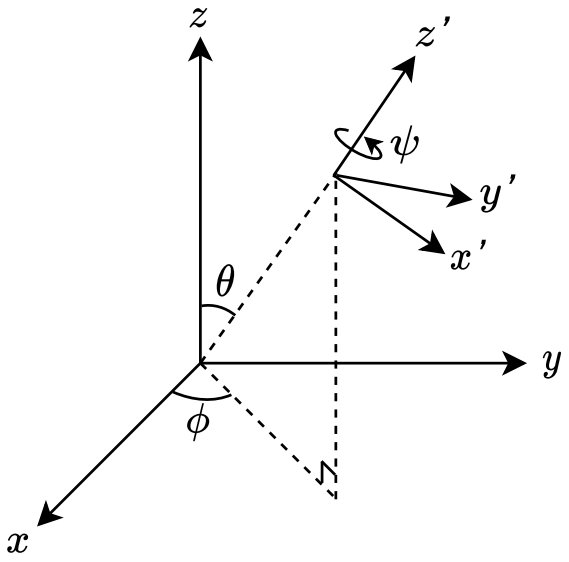


FIG. 2.3. Relative orientation of the source and sky frames.

$$h_{11} = h_+(\cos^2\theta\cos^2\phi\cos 2\psi - \sin^2\phi\cos 2\phi - \cos\theta\sin 2\phi\sin 2\psi) + h_\times(\cos^2\theta\cos^2\phi\sin 2\psi + \cos\theta\sin 2\phi\cos 2\psi - \sin^2\phi\sin 2\psi), \tag{2.88}$$

$$h_{22} = h_+(\cos^2\theta\cos 2\phi\sin^2\psi - \cos 2\phi\cos^2\psi + \cos\theta\sin 2\phi\sin 2\psi) + h_\times(\cos^2\theta\sin 2\phi\sin^2\psi - \cos\theta\cos^2\phi\sin 2\psi - \sin 2\phi\cos^2\psi). \tag{2.89}$$

Substituting this into $h = (h_{11} - h_{22})/2$ leads to the expression

$$h = h_+\left(\frac{1+\cos^2\theta}{2}\cos 2\phi\cos 2\psi - \cos\theta\sin 2\phi\sin 2\psi\right) + h_\times\left(\frac{1+\cos^2\theta}{2}\cos 2\phi\sin 2\psi + \cos\theta\sin 2\phi\cos 2\psi\right). \tag{2.90}$$

Therefore, the phase difference induced by the GW is expressed as

$$\Delta\phi_{\rm GW} = \Omega_0\int_{t-2L/c}^{t} \mathrm{d}t'\left(h_+(t')F_+(\theta,\phi,\psi) + h_\times(t')F_\times(\theta,\phi,\psi)\right), \tag{2.91}$$

where F_+ and $F_\times$ are called the *antenna pattern functions*, given by

$$F_+(\theta,\phi,\psi) = \frac{1+\cos^2\theta}{2}\cos 2\phi\cos 2\psi - \cos\theta\sin 2\phi\sin 2\psi, \tag{2.92}$$

$$F_\times(\theta,\phi,\psi) = \frac{1+\cos^2\theta}{2}\cos 2\phi\sin 2\psi + \cos\theta\sin 2\phi\cos 2\psi, \tag{2.93}$$

respectively. These functions represent the sensitivity of the Michelson interferometer to the plus and cross polarizations of GWs coming from the direction (θ,ϕ,ψ). Figure 2.4 illustrates the antenna patterns of a Michelson interferometer in the $z = 0$ plane with arms along the x and y axes for both plus and cross modes, along with the average.

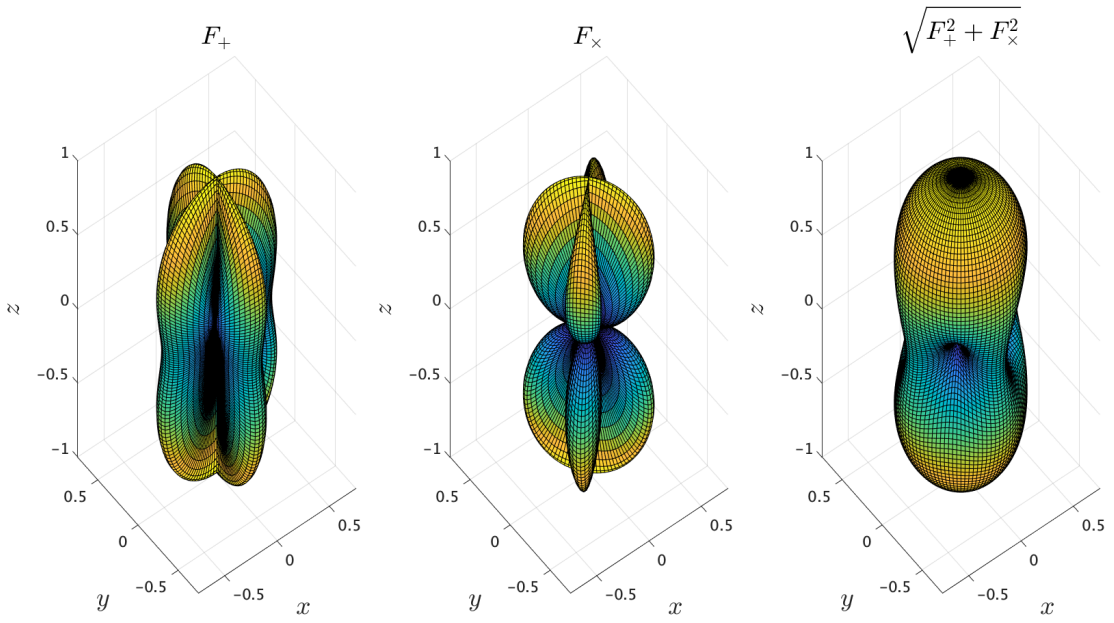


FIG. 2.4. Antenna pattern of a Michelson interferometer in the $z = 0$ plane with arms along the x and y axes. The polarization angle ψ is set to zero.

2.2.3 Current detectors

The GW detectors currently in operation are second-generation detectors such as Advanced LIGO, Advanced Virgo, and KAGRA. GEO600 is also in operation to collect data through the *astrowatch* program. GEO600 also plays an important role in demonstrating novel technologies for other detectors.

Configuration of current detectors

The standard configuration of the current detectors is called the dual-recycled Fabry-Pérot Michelson interferometer. This setup comprises several essential components to enhance the sensitivity of GW detection. Figure 2.5 shows a schematic representation of this configuration. There are three primary techniques integrated in a simple Michelson interferometer.

- **Fabry-Pérot cavity.** The Fabry-Pérot cavity consists of two highly reflective mirrors arranged in parallel. It works as an optical resonator, allowing light to bounce back and forth between the mirrors multiple times, increasing the effective path length traveled by the laser light.

- **Power recycling.** Since the interferometer is controlled to be at the dark fringe, most of the incident light is returned to the laser. In the power recycling technique, a mirror is placed at the injection port of the interferometer and reflects the returned light back to the interferometer. The power recycling cavity, which consists of the power recycling mirror and the input test masses, effectively increases the injection laser power, thereby reducing shot noise.

- **Signal recycling.** The signal recycling mirror, placed at the anti-symmetric port of the interferometer, reflects the signal sidebands into the interferometer, forming the

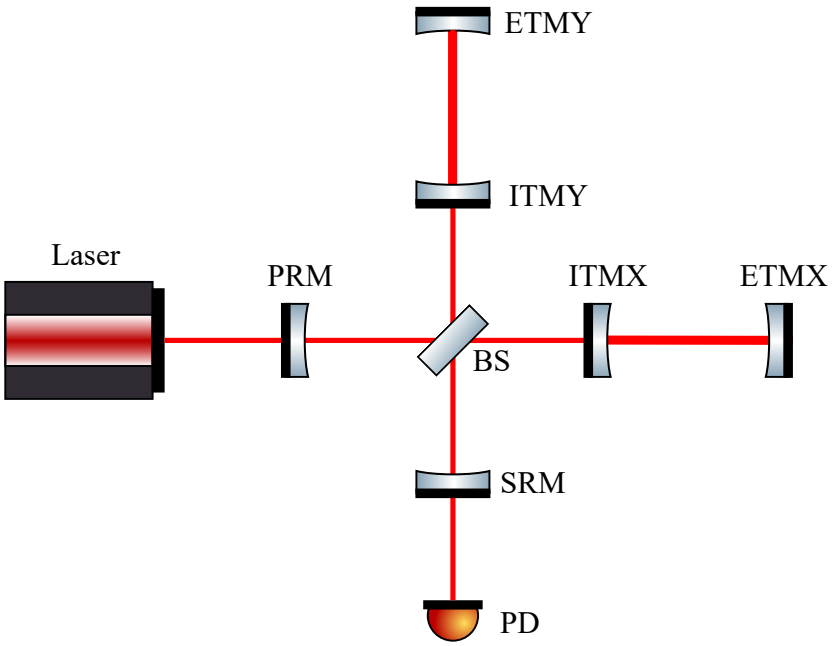


FIG. 2.5. Schematic diagram of a dual recycled Fabry-Pérot Michelson interferometer. ITMX (Y): input test mass X (Y), ETMX (Y): end test mass X (Y), BS: beam splitter, PRM: power recycling mirror, SRM: signal recycling mirror, PD: photo detector.

signal recycling cavity with the input test masses. If the signal recycling mirror is placed such that the GW sidebands are resonant within the cavity, the technique is called *resonant sideband extraction*, allowing extraction of the GW sidebands before they are canceled out. When the signal recycling cavity is detuned, sensitivity is improved in a narrow frequency range.

Noise sources

The main sources of noise in current ground-based interferometers are seismic noise, thermal noise, and quantum noise. Figure 2.6 shows the design sensitivity and noise spectrum of KAGRA.

- **Seismic noise.** The detector's mirrors are sensitive to various disturbances such as micro-seismic motion, wind, and human activity, which are attenuated by suspension pendulums. Seismic noise from ground vibrations limits the sensitivity below $\sim$ 10 Hz, where the strain sensitivity is inversely proportional to the square of the frequency [60]. A single pendulum with resonance frequency f_0, at frequencies $f \gg f_0$ attenuates the strain sensitivity by a factor of f_0^2/f^2. To further improve vibration isolation, multi-stage pendulums are used in current detectors. In addition, the construction of detectors underground, such as KAGRA, further reduces the seismic noise.

- **Thermal noise.** Thermal noise arises from random energy transfer with the heat bath, inducing vibrations in both the suspension system and the mirror. Suspension thermal noise results from the mechanical vibrations generated within the suspension system, inducing small fluctuations in the mirror's position. Mirror thermal

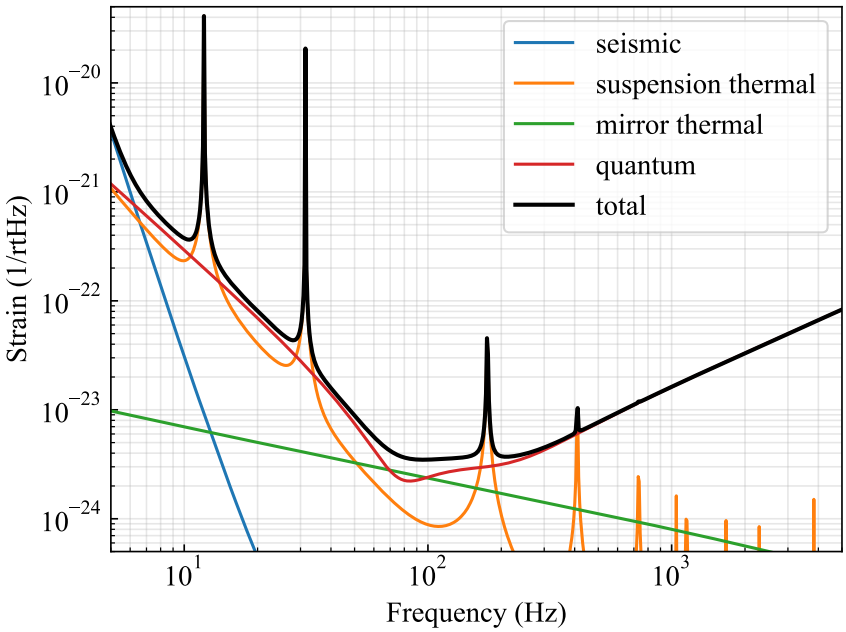


FIG. 2.6. Design sensitivity of KAGRA [59].

noise involves the excitation of the elastic vibrations of the mirror, resulting in tiny oscillations on its surface. Typically, suspension thermal noise is the dominant noise source in the frequency range of 10 Hz to 100 Hz. Around 100 Hz, mirror thermal noise becomes prominent. Methods to reduce thermal noise include the use of high-Q coatings and suspension wires, and cooling the mirrors and suspension wires. Specifically, KAGRA is designed to lower the temperature of its mirrors to 20 K [61].

- **Quantum noise.** Quantum noise originates from vacuum field fluctuations entering the anti-symmetric port of the interferometer. At high frequencies, shot noise due to the phase fluctuations of the vacuum field is dominant, while at low frequencies, radiation pressure noise caused by the amplitude fluctuations is prominent. There exists a trade-off relationship between these two, resulting in a lower bound on the noise spectrum, known as the *standard quantum limit* [62]. Utilizing squeezed light can be used to surpass the standard quantum limit [63].

2.2.4 Future detectors

Various types of interferometric detectors are planned to be constructed in the future to advance GW astronomy.

Ground-based detectors

LIGO-India [64], an almost identical replica of the current LIGO detectors, is planned for construction in Hingoli, India. This will improve the precision of GW source localization and increase the operational time, also known as the *duty cycle*, of the international GW network. Neutron-star Extreme Matter Observatory [65] is another proposed future detector in Australia. By using long signal recycling cavity, the detector specializes in

the 1-4 kHz frequency band, making it ideal for detecting the post-merger phase of BNS mergers and CCSNe to study nuclear physics.

Following the success of the second-generation detectors, third-generation detectors have been proposed and will be constructed in a few decades. Einstein Telescope [66] is a detector proposed to be built underground in Europe. Although its optical layout has not been finalized, one concept is to arrange three detectors in a triangular configuration, each consisting of two interferometers with 10-km long arms, shown in Fig. 2.7. The two interferometers are specialized for low-frequency and high-frequency bands, respectively [67]. Cosmic Explorer [68] is another proposed third-generation detector in the United States, consisting of an interferometer with 40-km L-shaped arms. These third-generation detectors are expected to make it possible to tackle a lot of interesting scientific questions [69, 70]. For instance, observing NSBH and BNS mergers with high signal-to-noise ratio (SNR) would provide valuable insights into the interior structure of neutron stars. Additionally, it could enable the first-ever detection of GWs from CCSNe and continuous waves from isolated rotating neutron stars.

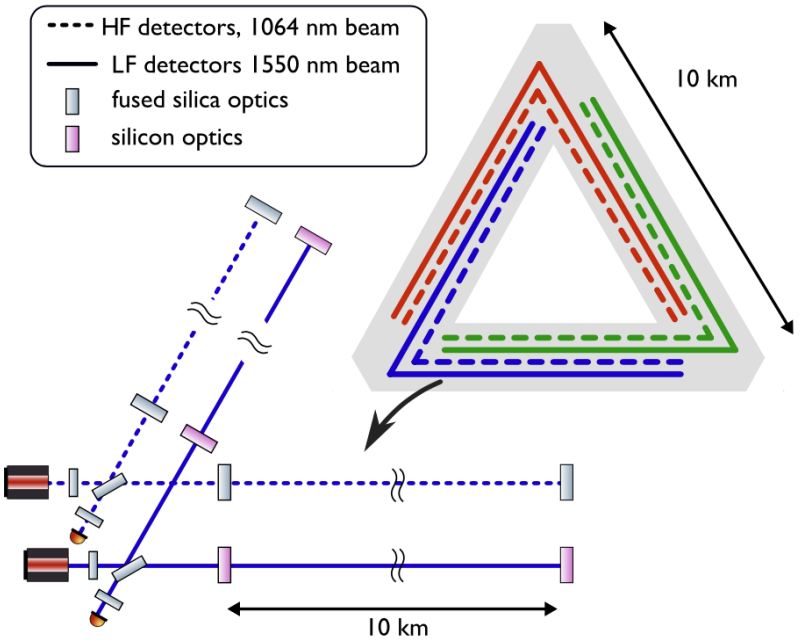


FIG. 2.7. Schematic illustration of the Einstein Telescope layout. Three detectors form an equilateral triangle, each consisting of two interferometers with a length of 10 km. One is optimized for high frequencies, using a 1064 nm laser beam and fused silica optics, while the other, designed for low frequencies, uses a 1550 nm laser beam and silicon optics. Image taken from Ref. [71].

Space-based detectors

The sensitivities of ground-based detectors at low frequencies is limited by seismic noise. To overcome this, several projects are underway to launch space-based interferometers. Laser Interferometer Space Antenna (LISA) [72] aims to observe GWs in the

frequency range between 0.1 mHz and 1 Hz with a constellation of three spacecrafts arranged in an equilateral triangle with sides 2.5 million km long, flying in a heliocentric orbit. DECi-hertz Interferometer Gravitational-wave Observatory (DECIGO) [73] aims to observe GWs in the frequency range of 0.1 Hz to 10 Hz. Its sensitivity bridges the gap between the sensitivity of LISA and that of ground-based detectors. Other projects, such as TianQin [74] and Taiji [75], are also in development. These space-based detectors could detect GWs from supermassive black hole mergers, extreme mass ratio inspirals, and stochastic GW backgrounds. Figure 2.8 plots characteristic strain for various detectors and sources.

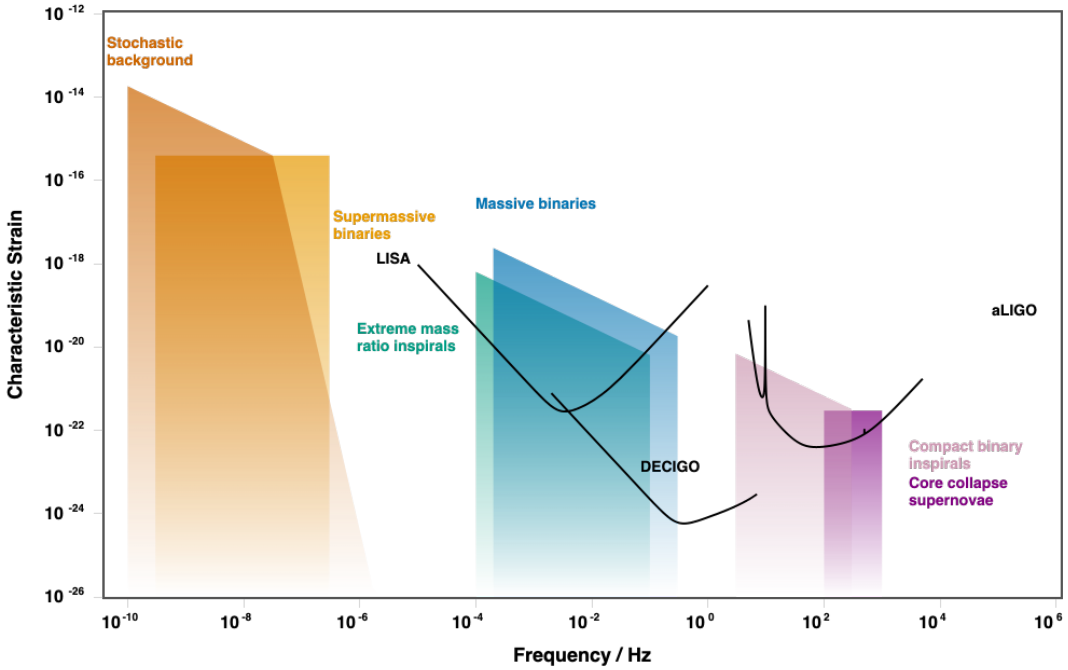


FIG. 2.8. Characteristic strain against frequency for various detectors and sources. The plot was generated using GWPlotter [76, 77].

2.3 Astrophysical source

2.3.1 Compact binary coalescence

Black holes and neutron stars are both formed after the death of stars through gravitational collapse. When the mass of the core after the explosion exceeds a threshold known as the Tolman-Oppenheimer-Volkoff limit [78, 79], the gravitational force becomes so strong that it overwhelms all other forces, causing the core to collapse further. This collapse forms a singularity surrounded by an event horizon, which marks the boundary of the black hole. If the remaining core mass after an explosion is less than the threshold, the gravitational force is counteracted by the neutron degeneracy pressure, which prevents further collapse. The core then stabilizes as a neutron star.

These dense celestial objects often form binary systems that orbit around each other while emitting GWs, eventually merging. This phenomenon is known as compact binary

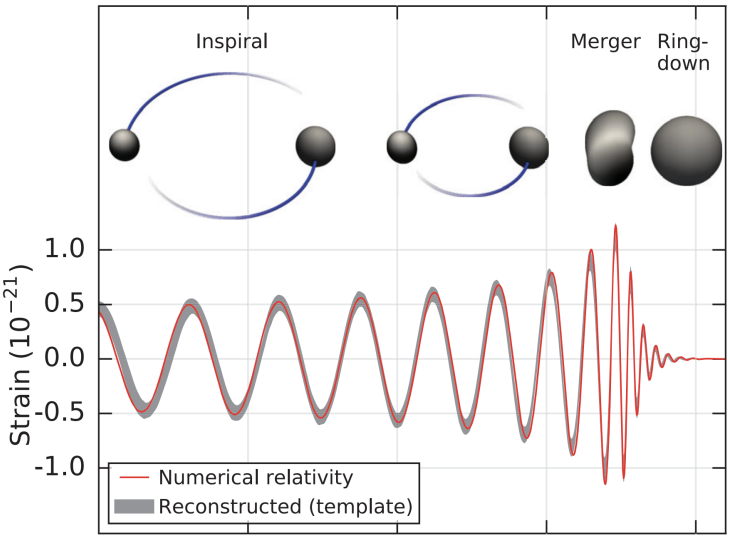


FIG. 2.9. Inpiral, merger, and ringdown phases of GW signals from CBC events. Image taken from Ref. [6].

coalescence (CBC). There are three distinct types of CBC events: BBH mergers, NSBH mergers, and BNS mergers. To date, all observed GWs have been produced by CBC events, with BBH mergers accounting for most of them. These GW signals exhibit three different signal morphologies: *inspiral*, *merger*, and *ringdown*. During the inspiral phase, the binary objects gradually approach each other and emit GWs of increasing frequency. Then, in the merger phase, the objects merge into a single object, emitting a burst of GWs. After the merger, the resulting object settles into a stable state, emitting damped oscillations in GWs as it reaches its final state. Figure 2.9 shows the inspiral, merger, and ringdown GW signal of a BBH merger.

Describing GWs from CBC sources requires at least 15 parameters, as summarized in Table 2.1. Additional parameters may include the eccentricity (e) of the binary orbit and the neutron star's tidal deformability (Λ). These 15 parameters are categorized as either intrinsic or extrinsic. Intrinsic parameters are the component masses and three-dimensional spin vectors. The dimensionless spin is defined as $\boldsymbol{\chi} = \boldsymbol{S}/m^2$, where $\boldsymbol{S}$ and m are the angular momentum and the mass of the component star, respectively. Extrinsic parameters are the spherical coordinates (r, ι, ψ), sky position, described by the declination α and the right ascension δ, along with the time and the phase at the coalescence, represented as (t_c, Φ_c).

Several techniques have been developed to accurately model GW waveform originating from CBCs. These waveforms are used for event detection and parameter estimation. The post-Newtonian approximation [80, 81] is a perturbative method to analytically solve the Einstein's field equations by expanding the system around v/c, where v represents the velocity of the objects. However, in binary systems, towards the end of the inspiral phase, the velocities can approach values close to the speed of light, making this method inaccurate for describing the entire signal. Another method is numerical relativity [82–84], which involves solving the Einstein's field equations numerically. This method is suited for accurately modeling the entire signal including the merger and the ringdown phases,

TABLE. 2.1. Parameters for describing GWs from CBC sources.

Intrinsic parameters	Component masses	m_1, m_2
	Component spins	$\boldsymbol{\chi}_1, \boldsymbol{\chi}_2$
Extrinsic parameters	Coalescence phase	Φ_c
	Coalescence time	t_c
	Luminosity distance	r
	Inclination	ι
	Polarization angle	ψ
	Declination	α
	Right ascension	δ

but it is highly computationally expensive. The effective-one-body formalism [85, 86] is a framework used to approximate the dynamics of binary systems with a single effective test mass, providing a full waveform. It is widely used by multiple detection pipelines due to its accuracy [34]. Phenomenological models [87, 88] combine the post-Newtonian approximations and the numerical relativity waveforms to generate a full waveform. These models are widely used for parameter estimation with their accuracy and computational efficiency [34].

As of January 2024, the total number of detected GW events from CBC sources is 171. Figure 2.10 displays the cumulative number of detections made from the O1 run to the middle of the O4 run.

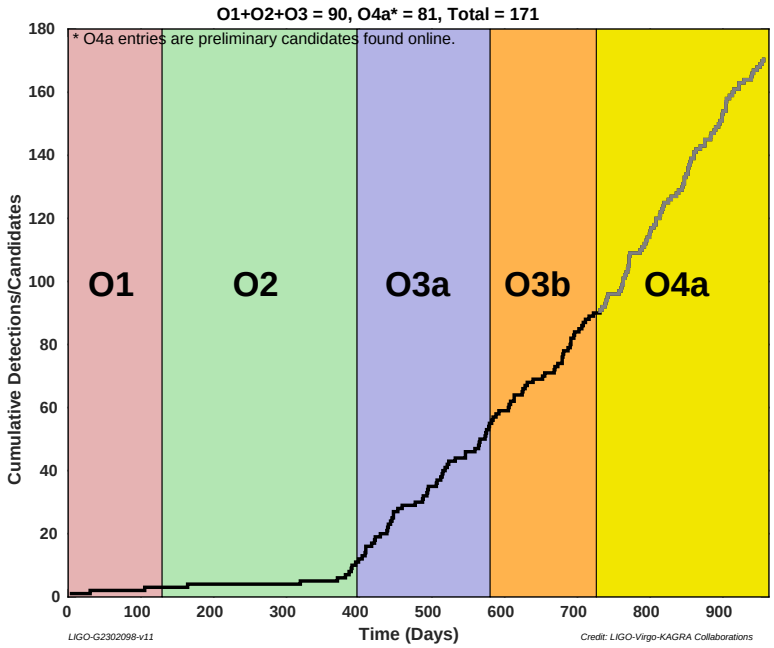


FIG. 2.10. Cumulative number of detections made from the O1 run to the middle of the O4 run. 'O3a' and 'O3b' refer to the first and second halves of the O3 run, respectively. Image taken from Ref. [89].

2.3.2 Core-collapse supernova

A supernova is an energetic explosion that marks the end of the life of a massive star, with an associated energy of $\sim 10^{53}$ erg across the electromagnetic spectrum, neutrinos, and GWs. Below we describe the basic theory of CCSNe, referring to review papers [90–92].

Throughout their lives, stars undergo nuclear fusion reactions in their interiors, creating a pressure that counteracts gravity. Initially formed from hydrogen and helium, stars are in hydrostatic equilibrium, but the energy loss through thermal radiation causes them to undergo a gravitational contraction, leading to an increase in density and temperature. Once the temperature exceeds 10^7 K, a hydrogen fusion reaction takes place in the star's core, marking its status as a main-sequence star. Subsequently, heavier elements such as carbon and oxygen are synthesized in sequence, forming an onion-skin structure with heavier elements lying toward the center. In light stars, nuclear fusion ceases with the synthesis of light elements, resulting in the formation of a white dwarf. Although the temperature decreases with the cessation of nuclear fusion, the white dwarf remains in equilibrium with the electron degeneracy pressure. Conversely, massive stars with masses above 8-10 $M_\odot$ continue the fusion process until iron is synthesised in the core. Since iron has a high binding energy per nucleon, it cannot undergo fusion. As iron begins to accumulate in the core, its mass increases. When the mass of the iron core exceeds the Chandrasekhar mass, gravity overcomes pressure and collapse begins. The resulting explosion is a CCSN explosion.

The detailed mechanism of the explosion in CCSNe has not been fully elucidated. The process is generally considered as follows.

- **Core collapse.** In the iron core, iron captures electrons through the reaction:

$$^{56}\mathrm{Fe} + \mathrm{e}^- \rightarrow {}^{56}\mathrm{Mn} + \nu_\mathrm{e}. \tag{2.94}$$

  This process decreases the number of electrons, lowering denegeracy pressure. Additionally, an endothermic reaction caused by iron photodisintegration:

$$^{56}\mathrm{Fe} + \gamma \rightarrow 13\,^4\mathrm{He} + 4n - 124.4\mathrm{MeV} \tag{2.95}$$

  further lowers the pressure and triggers core collapse.

- **Neutrino trapping.** After the collapse, the density of the core increases, and various reactions produce neutrinos. When the core density exceeds 10^{11}-10^{12} g/cm^3, electron neutrinos begin to be trapped in the core. This is called the *neutrino trapping*, and the region where neutrinos are trapped is called the *neutrino sphere*.

- **Core bounce.** During the collapse, the core consists of two parts. The inner core contracts at a subsonic speed while the outer core contracts at a supersonic speed. When nuclear densities reach $\sim 3 \times 10^{14}$ g/cm^3, the falling material bounces off the nucleus and creates a shock wave outward. The shock wave is initially generated spherically symmetrically inside the neutrino sphere. In the area where the shock wave passes through, the reaction $\mathrm{e}^- + p \leftrightarrow n + \nu_\mathrm{e}$ produces a large amount of

electron neutrinos. When the shock wave reaches the neutrino sphere, the neutrinos are emitted. This is called the *neutronization burst*. The core that was not shocked during the core bounce forms a proto-neutron star.

- **Shock stall.** As the shock wave propagates, it loses energy due to iron photodisintegration and neutrino emissions. The shock wave stops before reaching the surface of the star. This stopped shock wave is called a standing accretion shock. The radius of the shock wave is 100-200 km.

- **Shock revival.** How the stalled shock wave is revived to reach the stellar surface is not yet fully understood and has been actively studied using numerical simulations. Two primary mechanisms are currently favored. The first is called the *neutrino-driven mechanism*. The massive gravitational energy from the collapse is stored in the proto-neutron star and is brought out by the neutrinos. In this scenario, a part of the energy carried by neutrinos is used to revive the shock wave. Recent multi-dimensional simulations have revealed that convection [93] and the standing-accretion shock instability (SASI) [94, 95] also play crucial roles in shock revival.

  The second mechanism is known as the *magnetorotational mechanism*, in which the rotational energy of the star amplifies the magnetic field, and the magnetic energy contributes to the revival of the shock wave. Since not many stars rotate rapidly, most CCSNe are believed to be caused by the neutrino-driven mechanism rather than by this mechanism.

The GW signals from CCSNe are dominated by the oscillations of the proto-neutron star, known as the g-modes. In neutrino-driven explosions, numerical simulations have also observed low-frequency modes originating from the SASI. Figure 2.11 shows the two GW signals from three-dimensional numerical simulations by Kuroda *et al.* [95]. In both signals, there is a mode where the frequency increases with time from 100 Hz to 800-1000 Hz. In addition, we can see a low frequency mode around 100-200 Hz in the simulation that used the SFHx equation of state.

2.3.3 Rotating neutron star

Neutron stars often manifest as pulsars, which emit beams of electromagnetic radiation. Many pulsars rotate at high speeds, with frequencies of tens of Hz or higher. These stars emit GWs due to asymmetries around their rotation axis caused by the deformation of their outer layers, known as *crusts*, as well as fluid oscillations such as r-modes [96]. These GWs are emitted at a nearly constant frequency for a period longer than the observation time, and are called *continuous waves*. The typical amplitude of continuous waves is written as [97]

$$h \sim \frac{4\pi^2 G \epsilon I_{zz} f_{\rm GW}^2}{c^4 r}, \tag{2.96}$$

where r is the distance to the source, ϵ is the *equatorial ellipticity* defined as $\epsilon = |I_{xx} - I_{yy}|/I_{zz}$, and $f_{\rm GW}$ is the frequency of the radiated GW, which is twice the rotation frequency: $f_{\rm GW} = 2f_{\rm rot}$.

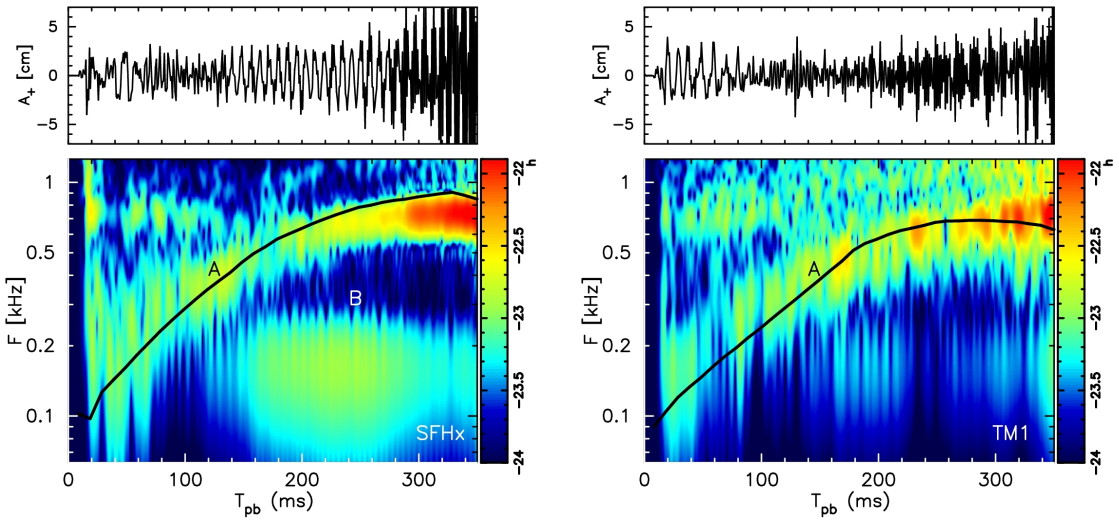


FIG. 2.11. Simulated GW amplitudes of plus mode and the corresponding time-frequency spectrograms by Kuroda *et al.* [95]. Equation of states are SFHx (left) and TM1 (right). The horizontal axis represents the time (ms) from core bounce. The simulation was stopped at 0.35 s after core bounce. The component 'A' is originated from the proto-neutron star's g-mode oscillation. The component 'B' is considered to be originated from the SASI. Figure taken from Ref. [95].

The frequency of continuous waves is nearly constant, but can change over the course of the observation due to energy loss from neutron stars and Doppler shift due to the Earth's revolution around the Sun. While ground-based detectors are sensitive within the frequency range of these waves, detecting them remains challenging due to their extremely small amplitude, with the estimated upper limit of strain being on the order of 10^{-25} [98]. Thus, to increase the SNR, we need to observe for a long time, such as several months or years, which is computationally expensive. Various methods, including those based on machine learning to search for continuous waves, are reviewed in Refs. [97, 99].

2.3.4 Stochastic background

The superposition of a large number of independent sources generates a stochastic GW background. These sources are divided into two categories: astrophysical and cosmological. Astrophysical sources, such as CBCs, CCSNe, rotating neutron stars, and magnetars, provide valuable information about the history of astrophysical processes. Cosmological sources, such as GWs from inflation, cosmic strings, and first-order phase transitions, could offer information about the early Universe [100].

The primary method to detect stochastic GW backgrounds is to take cross correlation of strain between two detectors. Let $s_i(t)$ be the ith detector output which is the sum of the detector noise $n_i(t)$ and GW signal $h_i(t)$:

$$s_i(t) = n_i(t) + h_i(t) \quad \text{for } i = 1, 2. \tag{2.97}$$

Time average of $s_1(t)s_2(t)$ is calculated as

$$\begin{aligned}\langle s_1(t)s_2(t)\rangle &= \langle n_1(t)+h_1(t)\rangle\langle n_2(t)+h_2(t)\rangle \\ &= \langle n_1(t)n_2(t)\rangle + \langle n_1(t)h_2(t)\rangle + \langle h_1(t)n_2(t)\rangle + \langle h_1(t)h_2(t)\rangle \\ &\sim \langle n_1(t)n_2(t)\rangle + \langle h_1(t)h_2(t)\rangle. \end{aligned} \tag{2.98}$$

If the noise at each detector is independent, only the GW term remains. However, even when two detectors are isolated, the detector noises are not completely independent. For instance, correlated magnetic noise from Schumann resonances should be taken into account [101, 102].

The intensity of stochastic GW backgrounds are usually characterized by the dimensionless energy density spectrum [103],

$$\Omega_{\rm GW}(f) = \frac{1}{\rho_c}\frac{{\rm d}\rho_{\rm GW}}{{\rm d}\log f}, \tag{2.99}$$

where $\rho_{\rm GW}$ is the energy density of the stochastic GW background, and $\rho_c = 3H_0^2/(8\pi G)$ is the critical energy density, and H_0 is the Hubble constant. Recent observations of binary mergers by Advanced LIGO and Advanced Virgo have provided insights into the stochastic GW backgrounds within the LIGO-Virgo observation band. The background energy density dominated by binary mergers is estimated to be $\Omega_{\rm GW}(f) \sim 10^{-9}$ at 25 Hz [104].

2.4 Data analysis

2.4.1 Matched filtering

Matched filtering is the optimal method for searching pre-known signals under stationary Gaussian noise, and it is currently the standard detection method for GWs from CBC sources. Here, we will explain the fundamentals of this method, following Ref. [105].

Gaussian noise and power spectral density

In a noise time series, if each sample is an independent Gaussian random variable with zero mean, the noise is known as Gaussian noise. The probability density function of a Gaussian noise $n(t)$ is written as

$$p[n(t)] \propto \exp\left[-\frac{1}{2}4\int_0^\infty {\rm d}f\ \frac{|\tilde{n}(f)|^2}{S_n(f)}\right], \tag{2.100}$$

where $S_n(f)$ is the one-sided power spectral density of the noise. For a time-series data $n(t)$ with an observation time T, the one-sided power spectral density is defined by

$$S_n(f) = \lim_{T\to\infty}\frac{2}{T}\int_0^\infty {\rm d}f\ |\tilde{n}(f)|^2. \tag{2.101}$$

If the power spectral density is independent of f, the noise is called *white noise*; otherwise it is called *colored noise*. By defining a noise-weighted inner product $(a|b)$ of two time

series $a(t)$ and $b(t)$ by

$$(a|b) = 4\mathrm{Re}\int_0^\infty \mathrm{d}f\ \frac{\tilde{a}(f)\tilde{b}^*(f)}{S(f)}, \tag{2.102}$$

the probability density function of a stationary Gaussian noise can be written as

$$p[n(t)] \propto e^{-(n|n)/2}. \tag{2.103}$$

We show an alternative expression for the power spectral density $S_n(f)$

$$\begin{aligned} S_n(f) &= \lim_{T\to\infty}\frac{2}{T}\int_{-T/2}^{T/2}\mathrm{d}t\ n(t)e^{i2\pi ft}\int_{-T/2}^{T/2}\mathrm{d}t'\ n(t')e^{-i2\pi ft'} \\ &= \lim_{T\to\infty}\frac{2}{T}\int_{-\infty}^{\infty}\mathrm{d}\tau\ e^{i2\pi f\tau}\int_{-T/2}^{T/2}\mathrm{d}t\ n(t)n(t+\tau) \\ &= 2\tilde{R}_n(f). \end{aligned} \tag{2.104}$$

Here, $\tilde{R}_n(f)$ is the Fourier transform of the auto-correlation function for $n(t)$, defined by

$$R_n(\tau) = \langle n(t)n(t+\tau)\rangle, \tag{2.105}$$

where $\langle\cdot\rangle$ denotes the temporal average. The expression in Eq. (2.104) is known as Wiener-Khinchin theorem. Another useful equation for $S_n(f)$ is as follows:

$$\begin{aligned} \langle\tilde{n}^*(f')\tilde{n}(f)\rangle &= \left\langle\int_{-\infty}^{\infty}\mathrm{d}t\ n(t)e^{-i2\pi f't}\int_{-\infty}^{\infty}\mathrm{d}t'\ n(t')e^{i2\pi ft'}\right\rangle \\ &= \int_{-\infty}^{\infty}\mathrm{d}t\ e^{i2\pi(f-f')t}\int_{-\infty}^{\infty}\mathrm{d}\tau\ \langle n(t)n(t+\tau)\rangle e^{-i2\pi f\tau} \\ &= \frac{1}{2}\delta(f-f')S_n(f). \end{aligned} \tag{2.106}$$

Optimal detection statistic

Suppose we aim to detect a GW signal whose waveform is known beforehand. Let us assume that the strain data $s(t)$ recorded by a detector comprises stationary Gaussian noise $n(t)$ and potentially a known GW signal characterized by parameters $\boldsymbol{\theta}$, denoted as $h(t;\boldsymbol{\theta})$. Detecting the signal $h(t;\boldsymbol{\theta})$ is considered a process of distinguishing between two hypotheses:

$$\mathcal{H}_0 : s(t) = n(t), \tag{2.107}$$

$$\mathcal{H}_1 : s(t) = h(t;\boldsymbol{\theta}) + n(t). \tag{2.108}$$

Here, $\mathcal{H}_0$ represents the null hypothesis, suggesting the absence of the GW signal in the data, while $\mathcal{H}_1$ represents the alternative hypothesis, indicating the presence of the GW signal $h(t;\boldsymbol{\theta})$ in the data.

According to the Neyman-Pearson lemma [106], the likelihood-ratio test is the uniformly most powerful test given a significance level α. Under the null hypothesis $\mathcal{H}_0$, s follows a Gaussian distribution. Therefore, following the calculations from the previous section, the likelihood is proportional to $e^{-(s|s)/2}$. Under the alternative hypothesis

$\mathcal{H}1$, $s-h$ also follows a Gaussian distribution, making the likelihood proportional to $e^{-(s-h|s-h)/2}$. Consequently, the likelihood ratio becomes

$$\Lambda = \frac{e^{-(s-h|s-h)/2}}{e^{-(s|s)/2}} = e^{(s|h)}e^{-(h|h)/2}. \tag{2.109}$$

As a result, the rejection region for the null hypothesis can be expressed using constants k and k' as

$$\{s \mid \Lambda \geq k\} \iff \{s \mid (s|h) \geq k'\}. \tag{2.110}$$

We call $(s|h)$ the *matched filter* and denote as m:

$$m = 4\mathrm{Re}\int_0^\infty \mathrm{d}f \ \frac{\tilde{s}(f)\tilde{h}^*(f)}{S_n(f)} \tag{2.111}$$

In practical scenarios, the parameters $\boldsymbol{\theta}$ describing the GW signal are unknown. Thus, the likelihood-ratio test is adapted to use

$$\hat{\Lambda} = \sup_{\boldsymbol{\theta}} \Lambda(s, h(\boldsymbol{\theta})). \tag{2.112}$$

Let us investigate the statistical properties of the matched filter. If there is no signal in the strain data, i.e., $s(t) = n(t)$, then the expected value of $m = (n|h)$ is zero, given $\langle n \rangle = 0$. The variance is calculated as

$$\begin{aligned}
\mathrm{Var}[m] &= \langle m^2 \rangle \\
&= 4\int_{-\infty}^{\infty} \mathrm{d}f \int_{-\infty}^{\infty} \mathrm{d}f' \frac{\langle \tilde{n}(f)\tilde{n}^*(f')\rangle \tilde{h}(f')\tilde{h}^*(f)}{S_n(|f|)S_n(|f'|)} \\
&= 4\int_{-\infty}^{\infty} \mathrm{d}f \int_{-\infty}^{\infty} \mathrm{d}f' \frac{\frac{1}{2}S_n(|f|)\delta(f-f')\tilde{h}(f')\tilde{h}^*(f)}{S_n(|f|)S_n(|f'|)} \\
&= 2\int_{-\infty}^{\infty} \mathrm{d}f \frac{\tilde{h}(f)\tilde{h}^*(f)}{S_n(|f|)} \\
&= (h|h).
\end{aligned} \tag{2.113}$$

Since $n(t)$ is a Gaussian random process and $m(t)$ is linear in $n(t)$, the matched filter $m(t)$ follows a Gaussian distribution with mean zero and variance $(h|h)$ when there is no signal. When the detector strain contains a GW signal $h(t)$, we obtain

$$\langle m \rangle = (h|h), \tag{2.114}$$

$$\mathrm{Var}[m] = (h|h), \tag{2.115}$$

thus the matched filter $m(t)$ follows a Gaussian distribution with mean $(h|h)$ and variance $(h|h)$. We define the SNR to be a normalized matched filter:

$$\rho = \frac{(s|h)}{\sqrt{(h|h)}}. \tag{2.116}$$

When the strain is pure noise, ρ follows a standard normal distribution $\mathcal{N}(0,1)$ and when the strain contains the signal $h(t)$, ρ follows a Gaussian distribution with mean $\sqrt{(h|h)}$

and variance one.

2.4.2 Parameter estimation

After a GW signal is detected, our primary objective is to extract and infer the values of the source parameters from the observed data. This inference process is essential for comprehensively studying the astrophysical properties of the detected source. In this context of parameter estimation, Bayesian statistics play a pivotal role, offering a methodological framework for quantifying parameter uncertainties through the GW posterior distribution. Bayes' theorem forms the bedrock of this statistical approach, enabling us to update our knowledge about parameters $\boldsymbol{\theta}$ given observed data d. Mathematically, Bayes' theorem is represented as

$$p(\boldsymbol{\theta}|d) = \frac{p(d|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(d)}. \tag{2.117}$$

Here, $p(\boldsymbol{\theta}|d)$ is the posterior probability distribution of parameters given the observed data. $p(d|\boldsymbol{\theta})$ is the likelihood function, expressing the probability of observing the data given the parameters. $p(\boldsymbol{\theta})$ is the prior distribution, representing our initial knowledge or assumptions about the parameters. $p(d)$ is the marginal likelihood or evidence, serving as a normalization factor. In most cases, calculating the posterior distribution $p(\boldsymbol{\theta}|d)$ analytically is infeasible due to high-dimensional and complex parameter spaces. This challenge is particularly evident in scenarios of the parameter estimation of GW signals from CBC sources, which are described by at least 15 parameters. A possible approach to estimate this complex distribution involves using computational methods that efficiently explore the parameter space and approximate the posterior distribution. Techniques such as Markov-chain Monte Carlo [107] and nested sampling [108] stand out as viable strategies for this purpose.

Markov-chain Monte Carlo

Markov-chain Monte Carlo (MCMC) methods, such as Metropolis-Hastings [107, 109] or Gibbs sampling [110], iteratively generate samples from the posterior distribution, aiming to converge towards an approximation of the true distribution. This method starts with an initial parameter values $\boldsymbol{\theta}^{(0)}$ and proposes new parameter values $\boldsymbol{\theta}'$ based on a proposal distribution $q(\boldsymbol{\theta}'|\boldsymbol{\theta}^{(t)})$. It then accepts or rejects these proposed values according to an acceptance probability $\alpha(\boldsymbol{\theta}^{(t)}, \boldsymbol{\theta}')$, which considers the ratio of the probabilities of the proposed and current values. Accepted values are used for the next iteration, and this process continues to generate a sequence of samples that approximate the target distribution.

Nested sampling

Nested sampling [108] is an efficient technique used to estimate the integral representing the evidence. This method aims to compute the multi-dimensional integral

$$p(d) = \int \mathrm{d}\boldsymbol{\theta}\ L(\boldsymbol{\theta})p(\boldsymbol{\theta}), \tag{2.118}$$

where $L(\boldsymbol{\theta}) = p(d|\boldsymbol{\theta})$ is the likelihood function. Nested sampling achieves the computation of this integral by transforming it into a one-dimensional integral using a series of steps. Initially, it involves employing a set of 'live points' within a likelihood iso-contours. Define $X(\lambda)$ as the integral over the parameter space limited to the regions where $L(\boldsymbol{\theta}) > \lambda$, weighted by the prior,

$$X(\lambda) = \int_{L(\boldsymbol{\theta})>\lambda} \mathrm{d}\boldsymbol{\theta}\ p(\boldsymbol{\theta}). \tag{2.119}$$

$X(\lambda)$ is a monotonically decreasing function of λ. Essentially, $X(\lambda)$ accumulates the prior mass within the iso-likelihood contours. Utilizing this, the evidence integral can be rewritten as

$$p(d) = \int_0^1 \mathrm{d}X\ L(X). \tag{2.120}$$

This transformation facilitates a computationally feasible approach to estimating the evidence by reducing the complexity associated with evaluating high-dimensional integrals.

2.4.3 Time-frequency analysis

Short-time Fourier transform

Short-time Fourier transform (STFT) is a method widely used to analyze the frequency evolution of a non-stationary signal over time by applying the Fourier transform to short segments of the signal. This is achieved by sliding a window function along the signal and computing the Fourier transform for each windowed segment. Using a window function w, the continuous STFT of a signal $x(t)$ is defined as

$$X(\tau, f) = \int_{-\infty}^{\infty} \mathrm{d}t\ x(t)w(t-\tau)e^{i2\pi ft}. \tag{2.121}$$

Its discrete form is given by

$$X[n, f] = \sum_{m=-\infty}^{\infty} x[m]w[m-n]e^{i2\pi fm}. \tag{2.122}$$

The square of the STFT is often used, called a spectrogram:

$$S(\tau, f) = |X(\tau, f)|^2. \tag{2.123}$$

The spectrogram provides an intuitive understanding of how the signal's spectrum changes over time. Usually it is plotted with time on the horizontal axis, frequency on the vertical axis, and color indicating energy density.

It can be shown that the duration σ_t and bandwidth σ_ω have an uncertainty relationship [111],

$$\sigma_t\sigma_\omega \geq \frac{1}{2}, \tag{2.124}$$

which will be derived in Appendix D. This relationship implies that a broader time observation provides higher-resolution frequency information and vice versa.

Wavelet transform

The Fourier transform is based on the idea that any signal can be decomposed into sinusoidal waves of different frequencies. Wavelet transform, on the other hand, uses an orthonormal basis generated by shifting and scaling a mother wavelet. The specific form of a wavelet scaled by a and translated by b is given by

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \tag{2.125}$$

where $\psi(t)$ is the mother wavelet. The continuous wavelet transform for a signal $x(t)$ is given by

$$W(a,b) = \int_{-\infty}^{\infty} \mathrm{d}t\ x(t)\psi^*_{a,b}(t). \tag{2.126}$$

While the STFT keeps the time and frequency resolutions constant, wavelet transform allows these resolutions to vary. The time duration is proportional to the parameter a, and the frequency spread is inversely proportional to a. These characteristics make wavelet transform useful for analyzing nonlinear and non-stationary signals.

Q transform

Q transform was initially introduced in Ref. [112] to analyze sounds from musical instruments. The method is now applied to many fields. For GWs, Chatterji *et al.* [113, 114] proposed to use it to detect burst signals, and it has become the standard technique to generate time-frequency representations of GW signals from CBC events. This method is a modification of the STFT with logarithmic tiling in frequency domain and a window function with a duration inversely proportional to the frequency. This modification provides a higher time resolution at higher frequencies.

For a time-series data $x(t)$, the conventional STFT can be expressed as

$$X(\tau, f) = \int_{-\infty}^{\infty} \mathrm{d}t\ x(t)w(t-\tau)e^{i2\pi ft}, \tag{2.127}$$

where $w(t-\tau)$ is a time-domain window function centered on time τ. Q transform introduces a window function with a duration inversely proportional to the frequency,

$$\sigma_t = \frac{C}{f}, \tag{2.128}$$

where C is a constant value. From the uncertainty relationship, the bandwidth is written as

$$\sigma_f = \frac{C'}{\sigma_t} = f\frac{C'}{C}, \tag{2.129}$$

where C' is another constant value. We define the quality factor Q with the ratio of central frequency to bandwidth,

$$Q = \frac{f}{\sigma_f} = \frac{C}{C'}. \tag{2.130}$$

By choosing a window function $w(t-\tau, f, Q)$ centered on time τ with a duration proportional to Q and inversely proportional to the frequency f, Q transform is defined by

$$X(\tau, f, Q) = \int_{-\infty}^{\infty} \mathrm{d}t\ x(t) w(t-\tau, f, Q) e^{i2\pi f t}. \tag{2.131}$$

This can be written in another way as

$$\begin{aligned} X(\tau, f, Q) &= \int_{-\infty}^{\infty} \mathrm{d}t \int_{-\infty}^{\infty} \mathrm{d}\phi \int_{-\infty}^{\infty} \mathrm{d}\phi'\ \tilde{x}(\phi') \tilde{w}^*(\phi, f, Q) e^{i2\pi(f+\phi-\phi')t} e^{-i2\pi\phi\tau} \\ &= \int_{-\infty}^{\infty} \mathrm{d}\phi \int_{-\infty}^{\infty} \mathrm{d}\phi'\ \tilde{x}(\phi') \tilde{w}^*(\phi, f, Q) \delta(f+\phi-\phi') e^{-i2\pi\phi\tau} \\ &= \int_{-\infty}^{\infty} \mathrm{d}\phi\ \tilde{x}(\phi+f) \tilde{w}^*(\phi, f, Q) e^{-i2\pi\phi\tau}. \end{aligned} \tag{2.132}$$

The standard window function in GW field is the Connes window [114], given by

$$\tilde{w}(\phi, f, Q) = \begin{cases} \left(\dfrac{315}{128\sqrt{11}} \dfrac{Q}{f}\right)^{1/2} \left[1 - \left(\dfrac{\phi Q}{f\sqrt{11}}\right)^2\right]^2 & \text{if } |\phi| < \dfrac{f\sqrt{11}}{Q}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.133}$$

The discrete Q transform is defined in the same way as the continuous version, as

$$X[m, l, Q] = \sum_{n=0}^{N-1} x[n] w[n-m, l, Q] e^{i2\pi n l/N}. \tag{2.134}$$

It can also be written as

$$X[m, l, Q] = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{x}[k+l] \tilde{w}^*[k, l, Q] e^{-i2\pi m k/N} \tag{2.135}$$

from Eq. (2.132). Therefore, the Q transform at a specific frequency is computed by performing a standard Fourier transform of the time series shifted in frequency, multiplying by a frequency-domain window function, and applying an inverse Fourier transform. In practice, Eq. (2.135) is more efficient than Eq. (2.134) in that the Fourier transform of the original time series only needs to be computed once. We then compute the inverse Fourier transform only for the logarithmically spaced frequencies of interest [113].

Figure 2.12 shows the Q-transformed time-frequency map of the GW150914 strain at the H1 detector. The plot shows the inspiral signal with increasing frequency.

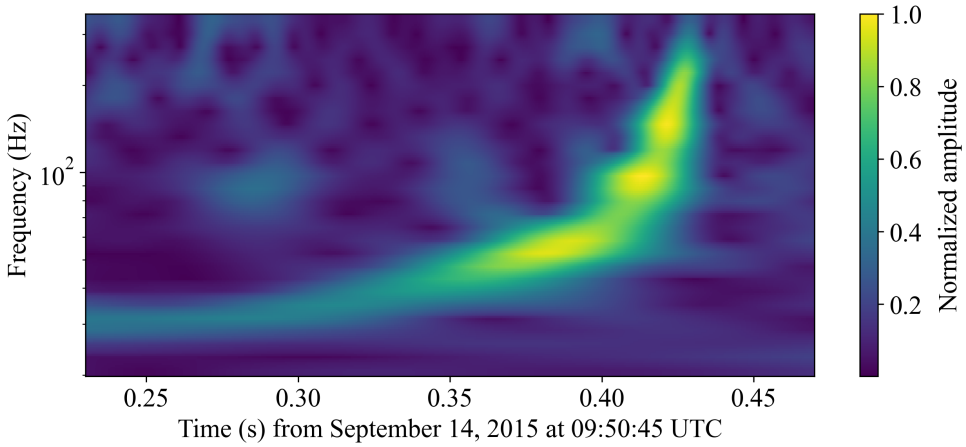


FIG. 2.12. Time-frequency representation of GW150914 at the H1 detector. The original strain time series is filtered with a 20-350 Hz band-pass filter and whitened.

Chapter 3

Deep Learning

This chapter introduces deep learning techniques which will be used in the research presented in Chapters 5 and 6. Particularly, we focus on explaining neural networks for supervised learning. We also review recent applications of machine learning to gravitational-wave data analysis.

3.1 Neural network

A neural network is a function that transforms input data into output data by passing it through a sequence of interconnected layers. In a neural network with l layers, the output x_l for an input x_0 can be expressed as

$$x_l = \left(f^{(l)} \circ f^{(l-1)} \circ \cdots \circ f^{(1)}\right)(x_0), \tag{3.1}$$

where $f^{(i)}$ represents the function applied by the ith layer of the network and $\circ$ represents the composite of functions. Various types of layers are commonly employed within neural networks.

3.1.1 Common layers

Fully connected layer

A *fully connected layer*, also known as a *dense layer*, is a fundamental component used in typical neural networks. Mathematically, it is defined by

$$f_{\text{dense}} : \mathbb{R}^{D_{\text{in}}} \to \mathbb{R}^{D_{\text{out}}};\ \boldsymbol{x} \mapsto \sigma(W\boldsymbol{x} + \boldsymbol{b}), \tag{3.2}$$

where $W \in \mathbb{R}^{D_{\text{out}} \times D_{\text{in}}}$ and $\boldsymbol{b} \in \mathbb{R}^{D_{\text{out}}}$ represent the weight matrix and bias vector, respectively. The function σ, referred to as the *activation function*, introduces non-linearity into the transformation process. One commonly used activation function is the rectified linear unit (ReLU) [115], defined as $\text{ReLU}(x) = \max(0, x)$.

Convolutional layer

A *convolutional layer* is a type of neural network layer that specializes in extracting features from input data, commonly used in image recognition tasks. In a two-dimensional convolutional layer, given a filter $K \in \mathbb{R}^{K \times K \times C_{\text{in}} \times N}$ and an input $X \in \mathbb{R}^{H_{\text{in}} \times W_{\text{in}} \times C_{\text{in}}}$,

the output of the layer with kernel size K, stride S, and padding P has the form $Y \in \mathbb{R}^{H_{\text{out}} \times W_{\text{out}} \times N}$. Here, H_{out} and W_{out} are expressed as

$$H_{\text{out}} = \left\lfloor \frac{H_{\text{in}} - K + 2P}{S} + 1 \right\rfloor, \quad W_{\text{out}} = \left\lfloor \frac{W_{\text{in}} - K + 2P}{S} + 1 \right\rfloor. \tag{3.3}$$

where $\lfloor \cdot \rfloor$ denotes the floor function. The (i, j, n) component of the output of the convolutional layer is calculated as

$$Y_{i,j,n} = \sum_{k=0}^{K-1} \sum_{l=0}^{K-1} \sum_{c=0}^{C_{\text{in}}-1} \tilde{X}_{iS+k,jS+l,c} K_{k,l,c,n}. \tag{3.4}$$

Here, $\tilde{X} \in \mathbb{R}^{(H_{\text{in}}+2P) \times (W_{\text{in}}+2P) \times C_{\text{in}}}$ represents the padded input, defined by

$$\tilde{X}_{i,j,c} = \begin{cases} X_{i-P,j-P,c} & \text{if } 0 \leq i - P < H_{\text{in}} \text{ and } 0 \leq j - P < W_{\text{in}}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.5}$$

Padding is often used to make the output size the same as the input.

Pooling layer

In a CNN, *pooling layers* are frequently used with convolutional layers. They play a role in reducing the spatial dimensions of the input data while retaining important information. Commonly, two types of pooling—*max pooling* and *average pooling*—are used. The output of a max-pooling layer with kernel size K, stride S, and padding P for an input $X \in \mathbb{R}^{H_{\text{in}} \times W_{\text{in}} \times C_{\text{in}}}$ is calculated as

$$Y_{i,j,n} = \max_{k=0,\ldots,K-1} \max_{l=0,\ldots,K-1} \tilde{X}_{iS+k,jS+l,n}, \tag{3.6}$$

while the output of an average-pooling layer is calculated as

$$Y_{i,j,n} = \frac{1}{K^2} \sum_{k=0}^{K-1} \sum_{l=0}^{K-1} \tilde{X}_{iS+k,jS+l,n}. \tag{3.7}$$

The padded input $\tilde{X}$ is defined in Eq. (3.5).

Dropout layer

A *dropout layer* [116] is a regularization technique used to prevent overfitting. During training, the dropout layer randomly sets a fraction p of input units to zero. This technique helps the network learn robust features. Mathematically, the dropout process involves generating binary masks,

$$m_1, \ldots m_{D_{\text{in}}} \text{ i.i.d.} \sim \text{Bernoulli}(1 - p) \tag{3.8}$$

and taking element-wise multiplication between the masks and the input,

$$\boldsymbol{y} = \boldsymbol{m} \cdot \boldsymbol{x}. \tag{3.9}$$

Batch normalization layer

In neural networks, *batch normalization* [117] is a technique that performs normalization for each mini-batch during training. A *mini-batch* refers to a subset of a data set that is processed together during one iteration of training. Instead of processing the entire dataset at once, which can be computationally expensive, the training data set is divided into smaller batches. These batches are randomly sampled from the dataset, and the model parameters are updated based on the gradients computed from each batch.

For an input comprised of n-dimensional vectors and a mini-batch size of m, denoted as $x \in \mathbb{R}^{m \times n}$, the output of the batch normalization layer is expressed as

$$y_{ij} = \gamma_j \frac{x_{ij} - \mu_j}{\sqrt{\sigma_j^2 + \epsilon}} + \beta_j, \tag{3.10}$$

where γ_j and β_j are trainable parameters of the layer. The mean μ_j and variance σ_j^2 for each dimension j of the input are computed as

$$\mu_j = \frac{1}{m} \sum_{i=1}^{m} x_{ij}, \tag{3.11}$$

$$\sigma_j^2 = \frac{1}{m} \sum_{i=1}^{m} (x_{ij} - \mu_j)^2. \tag{3.12}$$

ϵ is a small value to avoid division by zero.

3.1.2 Training process

Loss function

In the previous section, we defined neural networks and introduced various types of layers. In machine learning, the process of fixing the parameters within layers is called *training*. This process is performed iteratively using a training set, and predefined *loss function*, which measures the discrepancy between the predicted output of the network and the actual value for a given input. Suppose we have a data set $\{(X_1, y_1), \ldots, (X_m, y_m)\}$ and the predicted value for the ith input is $\hat{y}_i$. In a regression problem, mean squared error function, defined by

$$\mathcal{L} = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i)^2, \tag{3.13}$$

is commonly used as the loss function. In a classification task, cross-entropy loss function is favored. For C-class classification, the cross-entropy loss is mathematically defined as

$$\mathcal{L} = -\frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{C} y_i^{(j)} \log \hat{y}_i^{(j)}. \tag{3.14}$$

Here, the output y_i is expressed as a one-hot vector with a size of C, where $y_i^{(j)} = \delta_{jC_i}$. C_i denotes the class of ith sample.

The parameters in a network are updated many times by a chosen optimization algorithm to reduce the loss value. During this process, the same training set is used repeatedly. This number of iterations is called an *epoch*. However, if we rely on one training set to optimize the network parameters, the performance of the model on unknown data sets may become low. This problem is called *overfitting*. To avoid it, usually an additional data set, called *validation set*, is prepared. Throughout the training phase, the validation data set is used to evaluate the model's performance on unseen data by monitoring the loss value for the validation set.

Optimizer

Most of the optimization algorithms used in neural networks are based on the *gradient descent* method. In this method, the parameters $\boldsymbol{\theta}$ at the ith iteration is determined using the gradients of a loss function $\mathcal{L}$ computed with the current parameters:

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \eta \nabla \mathcal{L}\Big(\boldsymbol{\theta}^{(i)}\Big). \tag{3.15}$$

Here, η represents the predetermined parameter known as the *learning rate*, which controls how far the parameters are moved. Small learning rates lead to precise optimization but may require many steps to converge the loss function. On the other hand, large learning rates lead to fast optimization but have the risk of overshooting and missing the minima. Thus, tuning the learning rate is important to improve the model.

The crucial step in updating the model parameters via Eq. (3.15) is to compute gradients of the loss function with respect to the parameters of interest. *Backpropagation* is a fundamental technique in the training of neural networks that systematically computes gradients with respect to each parameter using the chain rule of derivatives. The chain rule is applied to compute gradients layer by layer, starting from the output layer and moving backward through the network.

In Eq. (3.15), since computing the loss function over the entire data set is computationally costly, a technique called *stochastic gradient descent* [118] has been developed. This method involves randomly sampling mini-batches from the original dataset to compute the loss function for the selected data and update the parameters based on the loss. The use of mini-batches is not unique to the stochastic gradient descent, but is a common practice in other optimization algorithms.

One method to avoid overfitting is *regularization*, in which a penalty term is added to the loss function when updating the parameters with the gradient descent. Two commonly used regularization techniques are L1 and L2 regularizations. For L1 regularization, the update equation in the gradient descent method is given by

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \eta \nabla \Big\{ \mathcal{L}\Big(\boldsymbol{\theta}^{(i)}\Big) + \lambda \Big\| \boldsymbol{\theta}^{(i)} \Big\|_1 \Big\}, \tag{3.16}$$

where the p-norm $\| \cdot \|_p$ is defined by

$$\|\boldsymbol{x}\|_p = \sum_i |x_i|^p. \tag{3.17}$$

For L2 regularization, the equation becomes

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \eta\nabla\Big\{\mathcal{L}\Big(\boldsymbol{\theta}^{(i)}\Big) + \lambda\Big\|\boldsymbol{\theta}^{(i)}\Big\|_2\Big\}. \tag{3.18}$$

In both regularizations, the hyperparameter λ controls the strength of regularization. While both L1 and L2 regularizations works to suppress overfitting, L2 regularization is preferred due to its easier computation of derivatives.

Several advanced optimization techniques have been developed to make the loss converged faster, which is outlined below. More comprehensive review of optimizers can be found in Ref. [119].

- **Momentum**. The stochastic gradient descent has difficulties when traveling across ravines—regions where the curvature is much more steeper in one dimension compared to others—that it oscillates across the slopes, resulting in slow convergence [119]. Momentum [120] helps accelerate the stochastic gradient descent by taking past updates into consideration. The updating equations are given by

$$\boldsymbol{v}^{(i)} = \gamma\boldsymbol{v}^{(i-1)} + \eta\nabla\mathcal{L}\Big(\boldsymbol{\theta}^{(i)}\Big), \tag{3.19}$$

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \boldsymbol{v}^{(i)}, \tag{3.20}$$

  where the hyperparameter γ is usually set to 0.9 or a similar value [119].

- **Adagrad**. In the stochastic gradient descent and the M omentum method, all parameters are updated with the same learning rate, but it is expected that using a different learning rate for each parameter can speed up convergence. Based on this idea, Adagrad [121] introduces a technique where the learning rate is adjusted using the magnitude of the gradients of each parameter. In this method, the parameters are updated by

$$\boldsymbol{g}^{(i)} = \nabla\mathcal{L}\Big(\boldsymbol{\theta}^{(i)}\Big), \tag{3.21}$$

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \frac{\eta}{\sqrt{\boldsymbol{G}^{(i)} + \epsilon}} \odot \boldsymbol{g}^{(i)}, \tag{3.22}$$

  where ϵ is a small value to avoid zero division, and $\odot$ denotes the Hadamard product. The jth component of $\boldsymbol{G}^{(i)}$ is defined as

$$G_j^{(i)} = \sum_{k\leq i}\Big(g_j^{(k)}\Big)^2. \tag{3.23}$$

- **RMSProp**. RMSProp [122] is another optimizer that incorporates an adaptive learning rate. The algorithm uses the following update equations:

$$\boldsymbol{g}^{(i)} = \nabla\mathcal{L}\Big(\boldsymbol{\theta}^{(i)}\Big), \tag{3.24}$$

$$\boldsymbol{v}^{(i)} = \beta\boldsymbol{v}^{(i-1)} + (1-\beta)\boldsymbol{g}^{(i)} \odot \boldsymbol{g}^{(i)}, \tag{3.25}$$

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \frac{\eta}{\sqrt{\boldsymbol{v}^{(i)} + \epsilon}} \odot \boldsymbol{g}^{(i)}. \tag{3.26}$$

The parameter β is a decay rate, commonly set around 0.9, controlling the contribution of the past gradients to the leaning rate.

- **Adam**. Adam [123] is one of the most popular algorithms today due to its stability. It incorporates both ideas from the momentum and the RMSProp. In this method, the updating equations are

$$\boldsymbol{g}^{(i)} = \nabla\mathcal{L}(\boldsymbol{\theta}^{(i)}), \tag{3.27}$$

$$\boldsymbol{m}^{(i)} = \beta_1\boldsymbol{m}^{(i-1)} + (1-\beta_1)\boldsymbol{g}^{(i)}, \tag{3.28}$$

$$\boldsymbol{v}^{(i)} = \beta_2\boldsymbol{v}^{(i-1)} + (1-\beta_2)\boldsymbol{g}^{(i)} \odot \boldsymbol{g}^{(i)}, \tag{3.29}$$

$$\hat{\boldsymbol{m}}^{(i)} = \frac{\boldsymbol{m}^{(i)}}{1-\beta_1^i}, \tag{3.30}$$

$$\hat{\boldsymbol{v}}^{(i)} = \frac{\boldsymbol{v}^{(i)}}{1-\beta_2^i}, \tag{3.31}$$

$$\boldsymbol{\theta}^{(i+1)} = \boldsymbol{\theta}^{(i)} - \frac{\eta}{\sqrt{\hat{\boldsymbol{v}}^{(i)}+\epsilon}} \odot \hat{\boldsymbol{m}}^{(i)}. \tag{3.32}$$

The values $\boldsymbol{m}^{(i)}$ and $\boldsymbol{v}^{(i)}$ are estimates of the first and the second moments. $\hat{\boldsymbol{m}}^{(i)}$ and $\hat{\boldsymbol{v}}^{(i)}$ are bias-corrected first and second moment estimates. The authors proposed the default parameters as $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$.

3.1.3 CNN architectures

Convolutional neural networks have shown promising performance, particularly in image recognition and computer vision tasks. They are designed to adaptively learn spatial features through convolutional and pooling layers, and based on extracted features, fully connected layers make predictions. Various CNN architectures have been proposed over the years. Among them, AlexNet [124] is one of the most influential CNN architectures in this field. It has revolutionized the field of deep learning by achieving overwhelming accuracy in the 1000-class classification task on the ImageNet dataset [125] using the deep architecture along with techniques such as data augmentation, dropout, ReLU layers, and the use of GPUs. VGGNet [126] is an improved version of the AlexNet with 16 or 19 layers. It is still used as a starting point for many image classification tasks.

These models demonstrated promising performance, but increasing the number of layers often led to the vanishing or exploding gradients problems. In response to this challenge, He *et al.* [127] proposed the residual neural network (ResNet), one of the most widely used CNN architectures today. The ResNet introduced several *residual blocks*, which consists of *skip connections*. The output of a residual block for an input x is expressed as

$$y = \sigma(F(x) + x), \tag{3.33}$$

where F represents the transformation by several layers, ensuring that the shape of the input remains unchanged. The skip connection adds the original input x to the transformed output $F(x)$ and an activation function σ is applied to the sum. During backpropagation,

the gradient of this block is calculated as

$$\frac{\partial y}{\partial x} = \frac{\partial y}{\partial H}\frac{\partial H}{\partial x} = \frac{\partial y}{\partial H}\left(\frac{\partial F}{\partial x} + 1\right), \tag{3.34}$$

where $H(x) = F(x) + x$. This equation indicates that even if gradients within the layers are small, the addition of the input allows for better gradient propagation. This concept has made it possible to train deep neural networks with tens to hundreds of layers. The specific architectures of ResNet are summarized in Table. 3.1.

TABLE. 3.1. ResNet architecures. The output sizes are shown for input with size 224×224. Taken from Ref. [127]

layer name	output size	34-layer	50-layer	101-layer
conv1	112×112	7×7, 64, stride 2		
conv2_x	56×56	3×3 max pool, stride 2		
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix} \times 4$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix} \times 23$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax		

3.2 Visualization technique

Machine learning predictions can often be opaque and difficult to comprehend, resembling a black box. In response to this challenge, explainable artificial intelligence has emerged as a prominent field. For CNNs, various visualization methods have been proposed to interpret the predictions of the models. Some of the most representative ones are introduced here.

3.2.1 Integrated gradients

Integrated gradients method [128] is a technique for computing the contributions of each input pixel to the predictions by approximating integrals from the output to the inputs along a straight line path. This algorithm is applicable to many neural networks.

Let F be a transformation of a neural network, x be the input, and x' be the baseline input. The feature attribution map is calculated by examining the path from the baseline x' to the input x and accumulating the network's gradients along this path. The integrated

gradients along the ith dimension for an input x are defined by

$$\mathrm{IG}_i(x) = (x_i - x_i') \int_0^1 \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i} \mathrm{d}\alpha. \tag{3.35}$$

3.2.2 Class activation mapping

Class activation mapping (CAM) [129] is a technique used to visualize and understand the regions of an input that contribute to the prediction of a particular class by a CNN. This method provides insights into the parts of the input that the model is focusing on when making its prediction.

CAM generates a map by computing the weighted combination of the feature maps of the last convolutional layer of the CNN. These feature maps capture the high-level representations of the input learned by the network. By examining the weights associated with each feature map, CAM can highlight the regions of the input image that are most relevant to the prediction of a specific class. The process of generating a CAM map is as follows:

1. Forward pass: The input is passed through the CNN, and the activations of the last convolutional layer are computed. These activations represent the high-level features learned by the network.

2. Global average pooling: The feature maps obtained from the last convolutional layer are aggregated using global average pooling. This step reduces the spatial dimensions of the feature maps to a single value per channel, effectively summarizing the information contained in each feature map.

3. Weighted combination: The aggregated feature maps are then combined with the weights to produce a class activation map. The resulting map highlights the regions of the input image that are most relevant for predicting the target class.

This procedure of calculating the CAM maps is illustrated in Fig. 3.1. By visualizing the class activation map overlaid on the input, one can gain insights into which parts of the image are important for the model's decision-making process.

3.2.3 Grad-CAM

Grad-CAM, an extension introduced by Selvaraju *et al.* [130], is a generalized version of the class activation mapping. Notably, the original CAM is exclusively applicable to networks that incorporate a global average pooling layer at the end. Grad-CAM overcomes this limitation by by extracting gradients from the final convolutional layer of the neural network and using them to generate a class-discriminative localization map.

Suppose that for a given input, the score for class c of the trained model is y^c, and the kth output matrix of the last convolutional layer is A^k. To obtain the Grad-CAM map of class c, we first compute the gradients of the score y^c with respect to the (i, j) component of the kth feature map A^k. We then take the global average of these gradients:

$$\alpha_k^c = \frac{1}{Z} \sum_{i,j} \frac{\partial y^c}{\partial A_{ij}^k}, \tag{3.36}$$

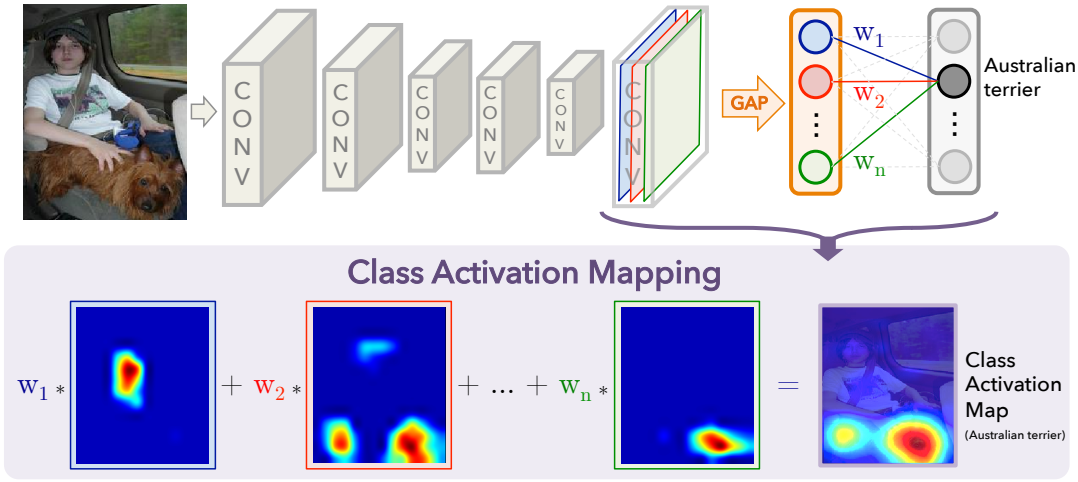


FIG. 3.1. Procedure for generating class activation maps. Image taken from Ref. [129].

where Z is the number of pixels in the feature map A^k. The Grad-CAM map of the class c is computed as linear sum of A^k with α_k^c as weights. The resulting map of class c is expressed as

$$L^c_{\text{Grad-CAM}} = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right). \tag{3.37}$$

The ReLU function is applied to extract only features that have a positive contribution to the score. Since convolutional layers and pooling layers make the size of the feature map smaller than the input, Grad-CAM map is finally interpolated to make it the same size as the input.

3.2.4 Grad-CAM++

While Grad-CAM takes a global average of the gradient matrix when calculating the weight α_k^c in Eq. (3.36), Chattopadhay *et al.* [131] proposed a method to fully include the importance of each pixel in the gradient matrix by taking its weighted average for the weight:

$$\alpha_k^c = \sum_{i,j} \alpha_{ij}^{kc}\ \text{ReLU}\left(\frac{\partial y^c}{\partial A_{ij}^k}\right). \tag{3.38}$$

The ReLU function is used to account for features that increase the activation of the output neuron rather than suppress the activation of the output neuron. The weights α_{ij}^{kc} can be derived as follows.

In a CNN with a global pooling layer, the final score y^c for a class c can be written as a linear combination of its feature maps A^k:

$$y^c = \sum_k \alpha_k^c \sum_{i,j} A_{ij}^k. \tag{3.39}$$

Combining Eqs. (3.38) and (3.39), we have

$$y^c = \sum_k \sum_{a,b} \alpha_{ab}^{kc} \, \mathrm{ReLU}\left(\frac{\partial y^c}{\partial A_{ab}^k}\right) \sum_{i,j} A_{ij}^k. \tag{3.40}$$

We take partial derivative with respect to A_{ij}^k on both sides of this equation. Without loss of generality, we can neglect the ReLU activation in the derivation as it only works as a threshold for allowing the gradients to flow back [131]. The first and second derivatives are calculated as

$$\frac{\partial y^c}{\partial A_{ij}^k} = \sum_{a,b} \alpha_{ab}^{kc} \frac{\partial y^c}{\partial A_{ab}^k} + \sum_{a,b} A_{ab}^k \alpha_{ij}^{kc} \frac{\partial^2 y^c}{(\partial A_{ij}^k)^2}, \tag{3.41}$$

$$\frac{\partial^2 y^c}{(\partial A_{ij}^k)^2} = 2\alpha_{ij}^{kc} \frac{\partial^2 y^c}{(\partial A_{ab}^k)^2} + \sum_{a,b} A_{ab}^k \alpha_{ij}^{kc} \frac{\partial^3 y^c}{(\partial A_{ij}^k)^3}, \tag{3.42}$$

respectively. From the second equation, we obtain

$$\alpha_{ij}^{kc} = \frac{\frac{\partial^2 y^c}{(\partial A_{ij}^k)^2}}{2\frac{\partial^2 y^c}{(\partial A_{ij}^k)^2} + \sum_{a,b} A_{ab}^k \frac{\partial^3 y^c}{(\partial A_{ij}^k)^3}}. \tag{3.43}$$

The CAM method using these weights is known as Grad-CAM++, since it can be considered as a generalization of Grad-CAM. The saliency map for Grad-CAM++ is expressed in the same way as for Grad-CAM, using weights in Eq. (3.38) and feature maps, as

$$L_{\mathrm{Grad-CAM++}}^c = \mathrm{ReLU}\left(\sum_k \alpha_k^c A^k\right). \tag{3.44}$$

3.2.5 Score-CAM

Wang *et al.* [132] proposed a gradient-free CAM method called Score-CAM. It solves the problem of gradient-based CAM methods, namely the gradient is unstable, easily disturbed by noise, and can vanish or explode in deep networks. To generate a Score-CAM map, feature maps are used to mask an input image. Let H^k be the k-th feature map, up-sampled to the same size as the input and normalized to [0, 1]. Given an input image X, the weight for the k-th feature map is computed as the difference between the score of the masked image $X \circ H^k$ and the score of the baseline image X_b:

$$\alpha_k = f(X \circ H^k) - f(X_\mathrm{b}), \tag{3.45}$$

where $f(\cdot)$ denotes the output of the CNN and $\circ$ denotes the Hadamard product. A black image is used as a baseline image. The Score-CAM map of class c is then computed as linear sum of the c-th value of α_k and the feature map A^k as

$$L_{\mathrm{Score-CAM}}^c = \mathrm{ReLU}\left(\sum_k \alpha_k^c A^k\right). \tag{3.46}$$

The flow of calculating the Score-CAM map is illustrated in Fig. 3.2.

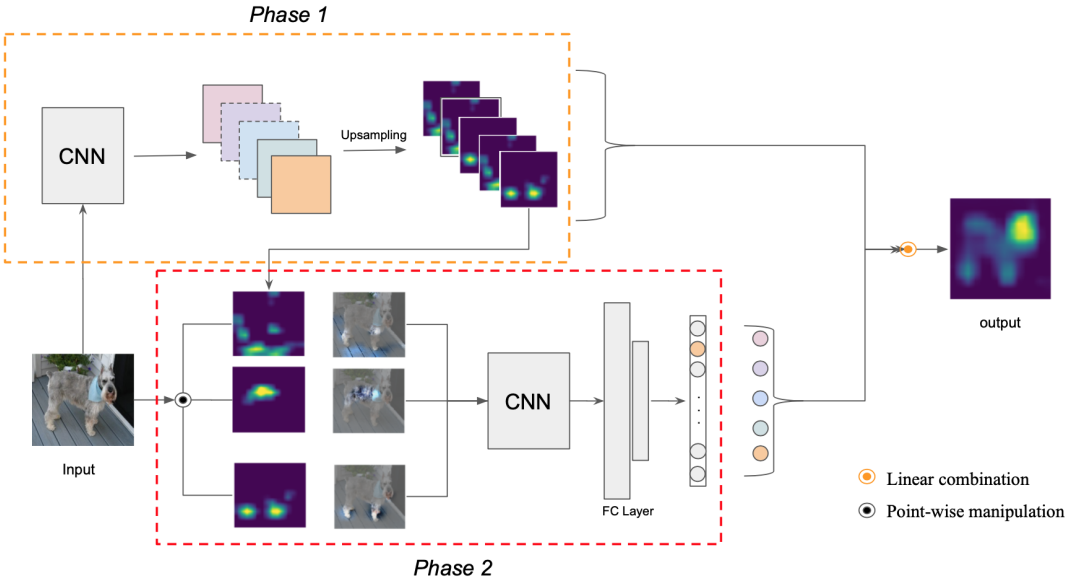


FIG. 3.2. Pipeline of Score-CAM. Image taken from Ref. [132].

3.2.6 Guided backpropagation

Guided Backpropagation [133] stands as another gradient-based visualization technique that visualizes the effect of each pixel of the input on the output of the model. The guided backpropagation map is computed by modifying the standard backpropagation algorithm to only propagate positive gradients while backpropagated through the ReLU functions. Let f_i^l be the ith output matrix of the lth layer which is followed by a ReLU layer:

$$f_i^{l+1} = \mathrm{ReLU}(f_i^l) = \max(f_i^l,\ 0). \tag{3.47}$$

In the normal backpropagation, the backpropagated value at layer l is calculated as

$$R_i^l = \frac{\partial f_i^{l+1}}{\partial f_i^l} \cdot \frac{\partial f^{\mathrm{out}}}{\partial f_i^{l+1}} \tag{3.48}$$

$$= \Theta(f_i^l) \cdot R_i^{l+1}, \tag{3.49}$$

where Θ is the Heaviside step function. On the other hand, guided backpropagation only allows the propagation of positive gradients:

$$R_i^l = \Theta(f_i^l) \cdot \Theta(R_i^{l+1}) \cdot R_i^{l+1}. \tag{3.50}$$

R^0 is the output image of the same size as input, which expresses the features of the input image that activate neurons in higher layers.

The authors of Ref. [130] proposed the Guided Grad-CAM technique, which can produce high-resolution maps. Grad-CAM maps are highly class-discriminative, but not high-resolution. In contrast, guided backpropagation maps are high-resolution, but not

class-discriminative [133]. The Guided Grad-CAM map takes advantage of both methods by taking element-wise multiplication of these two maps.

3.3 Machine learning in gravitational-wave astronomy

Following the remarkable development of machine learning in the fields of image recognition and natural language processing, the filed of physics, including GWs, has also seen the application of machine learning over the past decade. Recent applications of machine learning for GW data analysis will be outlined below. More comprehensive review can be found in Refs. [41, 42].

3.3.1 Signal detection

Various machine learning-based algorithms have been studied to detect different types of signals, including CBCs, CCSNe, continuous waves, and stochastic backgrounds.

Compact binary coalescence

Machine learning-based methods have been studied as an alternative to the traditional matched-filtering technique, which becomes computationally expensive when considering a large number of signal parameters. In a seminal study, George and Huerta [44] trained a one-dimensional CNN for distinguishing between pure noise and BBH signals. They demonstrated the CNN's ability to identify simulated BBH signals in Gaussian noise. Gabbard *et al.* [134] also performed a similar study around the same time. In Ref. [135], the authors pointed out that machine learning models are not suited to claim statistically significant detections of GW events, but can be useful for real time trigger generation. Some studies proposed models inspired by or combined with the matched-filtering techniques [136, 137]. Recently, a mock data challenge of machine learning-based BBH detection models was conducted [138] and compared with the traditional PyCBC pipeline [37, 139]. In their study, machine learning-based models showed comparable performance to the matched-filtering technique in Gaussian noise, but struggled against real detector noise. Subsequently, Nousi *et al.* [140] reported that their model outperformed the PyCBC pipeline even in the real detector noise of O3.

For the detection of BNS signals, Krastev *et al.* [45, 141] demonstrated CNN-based approaches to search for BNS and BHH signals in 10 sec real detector data. Their model recovered all BBH signals in GWTC-1 as well as the BNS event, GW170817. In Refs. [142, 143], only inspiral parts of BNS signals are input to CNNs to provide early warnings before mergers. Aveiro *et al.* [144] applied 'you only look once' (YOLO) object detection model [145] to detect BNS signals in a spectrogram.

There are fewer studies focusing on NSBH signals compared to BBH and BNS signals. In Refs. [146, 147], CNN models were built for detecting all signal types of CBC events, including NSBH mergers.

Core-collapse supernova

Recent multi-dimensional numerical simulations have enabled the training of machine-learning models for detecting CCSNe by using the simulated signals and insights gained from them. The first attempt was made by Astone *et al.* [46]. They used phenomenological signals of g-mode to train CNN models [46], surpassing the coherent WaveBurst pipeline [148, 149] in their dataset. This method was further updated in Ref. [47] by using advanced phenomenological waveforms and CNN models. In other studies [150–152], CNNs were trained using CCSN signals from various numerical simulations. Classification of equation of states using simulated magnetorotational signals were carried out in Refs. [153, 154]. Powell *et al.* [155] compared various methods including CNNs for classifying multiple explosion mechanisms.

Continuous wave

Detecting GWs from isolating neutron stars is a challenging task due to their small amplitude, which requires the processing of long duration data, resulting in high computational cost. In Ref. [156], the authors explored Doppler parameter estimation using a conditional variational autoencoder for rapid follow-up analyses. In another study by Yamamoto and Tanaka [157], a combination of an excess-power method and a CNN was employed for an all-sky search in $\mathcal{O}(10^7)$ sec-long data. End-to-end deep learning-based methods have been developed in Refs. [158–160].

Stochastic background

There are not so many studies that use machine learning to search for stochastic GW backgrounds. Among various candidates for stochastic backgrounds, search methods for those from stellar-mass BBH mergers have been investigated in Refs. [161, 162].

3.3.2 Parameter estimation and sky localization

Parameter estimation, including sky localization of GW sources, has relied on Bayesian techniques such as the Markov-chain Monte Carlo method. However, as the number of signal parameters increases, the computational cost becomes prohibitive. Typically, the full parameter estimation of CBC signals can take hours to days. Reducing the computational cost of sky localization is particularly important for NSBH and BNS signals to enable multi-messenger observations.

As an alternative to the current method, machine learning based methods have been explored to reduce the latency. Although regression and classification models can be applied to these tasks [1, 163], methods to estimate the posterior distribution are often preferred. Gabbard *et al.* [48] used a conditional variational autoencoder to estimate posterior distributions of BBH parameters. Another popular model for this purpose is the normalizing flow [164]. It works by transforming a simple initial distribution, such as a Gaussian, into a more complex distribution through a series of invertible mappings. The DINGO software [49, 50] has been developed using this method. Chatterjee *et al.* [165] also used a normalizing flow for fast sky localization not only for BBH signals but also

for NSBH and BNS signals. Additionally, they demonstrated the sky localization of BNS signals before their merger time [166]. These machine learning-based sky localization methods have the potential to improve multi-messenger astronomy in the future.

3.3.3 Glitch study

The detector output has many non-Gaussian and non-stationary noise transients, known as *glitches*. These are caused by instrumental and environmental disturbances from many different sources, some known and some yet to be identified. It is of great importance to detect and subtract these noise transients to reduce the false alarm rate in the search for GWs. Glitches can sometimes overlap with GW signals, making it difficult to accurately estimate signal parameters. A notable example of this occurred during the GW170817 event at the LIGO Livingston detector, where a loud glitch overlapped with the GW signal, as shown in Fig. 3.3. This example emphasizes the importance of identifying and mitigating glitches from detector data. It is also essential to know the frequency and nature of glitches to improve the sensitivity of detectors.

The Gravity Spy project [51, 53] is a noteworthy initiative in this field and has played a key role in studying glitches. In this project, spectrogram images of glitches identified by the Omicron software [167] are labelled with 22 different classes based on their source or shape (e.g. 'Blip', 'Koi Fish', and 'Scattered Light'). Various supervised learning models have been proposed to classify the glitches in this dataset. For instance, Bahaadini *et al.* [52] showcased the effectiveness of using merged images of different time durations to address the different time scales of glitches. Powell *et al.* [168] used a generative adversarial network [169] to produce fake glitches from the Gravity Spy dataset and trained a classifier of them, achieving 99% classification accuracy. Sakai *et al.* [170, 171] proposed an unsupervised learning model for clustering glitches. This not only reduced annotation costs, but also potentially revealed the existence of new classes not present in the Gravity Spy dataset.

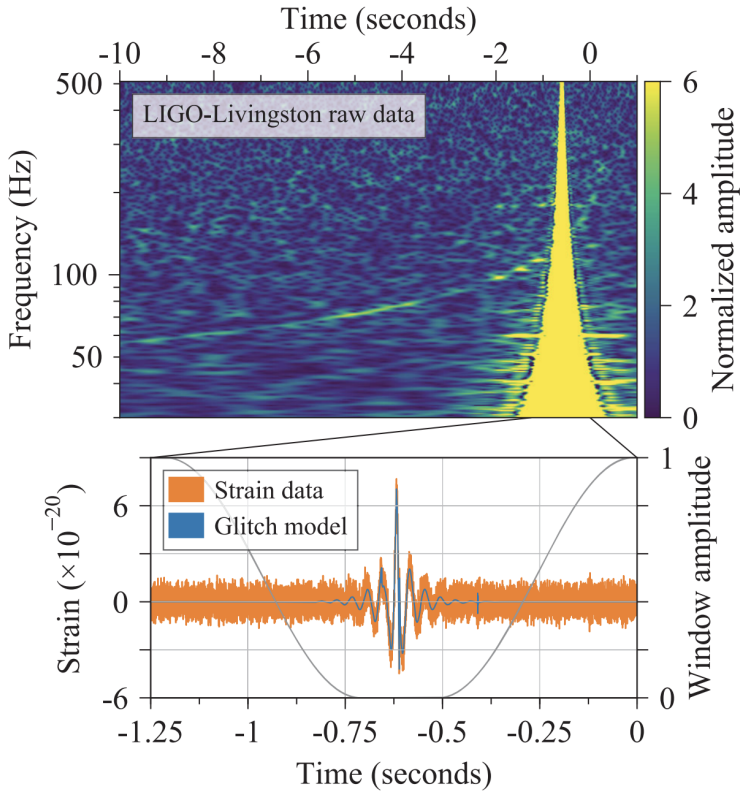


FIG. 3.3. The glitch overlapped with the GW170817 event in the LIGO Livingston data. Figure taken from Ref. [20].

Chapter 4

Hilbert-Huang Transform

The Hilbert-Huang transform [56] stands as a powerful signal processing technique for exploring non-stationary and nonlinear signals by providing high-resolution time-frequency representations. This chapter navigates through the fundamental aspects of the Hilbert-Huang transform, primarily following Refs. [56, 172].

4.1 Analytic signal

Signals are real in nature; however, by defining a complex signal that corresponds to the real signal, we can determine an instantaneous frequency of the signal. Here, we introduce the concept of an *analytic signal* and delve into its properties concerning the determination of instantaneous amplitude and instantaneous frequency.

Let $z(t)$ be a complex signal whose real part is the original signal $s(t)$, while the imaginary part $s_i(t)$ is determined in a specific way:

$$z(t) = s(t) + is_i(t) = a(t)e^{i\phi(t)}. \tag{4.1}$$

By fixing the imaginary part, we can define the instantaneous amplitude $a(t)$ and the instantaneous phase $\phi(t)$ as

$$a(t) = \sqrt{s(t)^2 + s_i(t)^2}, \tag{4.2}$$

$$\phi(t) = \tan^{-1}\left(\frac{s_i(t)}{s(t)}\right). \tag{4.3}$$

To define the imaginary part effectively, consider the conjugate symmetry of the spectrum for real signals,

$$\tilde{s}(-\omega) = \tilde{s}^*(\omega). \tag{4.4}$$

Since $|\tilde{s}(\omega)|^2$ is symmetric, the average angular frequency is always zero,

$$\langle\omega\rangle = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\ \omega|\tilde{s}(\omega)|^2 = 0, \tag{4.5}$$

which is not intuitive. To address this, we define a complex signal by considering only the positive frequencies. Thus, the Fourier transform of the complex signal is expressed as

$$\tilde{z}(\omega) = 2\tilde{s}(\omega)\Theta(\omega), \tag{4.6}$$

where Θ is the Heaviside step function. The factor 2 ensures that the real part is equal to the original signal. The inverse Fourier transform of this complex signal yields

$$\begin{aligned}
z(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\ \tilde{z}(\omega)e^{-i\omega t} \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\ \tilde{s}(\omega)\Theta(\omega)e^{-i\omega t} \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\ \tilde{s}(\omega)\frac{1}{2}(1+\mathrm{sgn}(\omega))e^{-i\omega t} \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\ \tilde{s}(\omega)e^{-i\omega t} + \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\ \tilde{s}(\omega)\mathrm{sgn}(\omega)e^{-i\omega t} \qquad (4.7)\\
&= s(t) + s(t) * \frac{i}{\pi t} \qquad (4.8)\\
&= s(t) + i\frac{1}{\pi}\mathrm{PV}\int_{-\infty}^{\infty} \mathrm{d}t'\ \frac{s(t')}{t-t'}, \qquad (4.9)
\end{aligned}$$

where $\mathrm{sgn}(\cdot)$ denotes the sign function, $*$ denotes the convolution, and PV denotes the Cauchy principal value. This imaginary part is known as the *Hilbert transform* of the original signal,

$$\mathcal{H}[s(t)] = \frac{1}{\pi}\mathrm{PV}\int_{-\infty}^{\infty} \mathrm{d}t'\ \frac{s(t')}{t-t'}. \qquad (4.10)$$

Therefore, the analytic signal can be written as

$$\mathcal{A}[s(t)] = s(t) + i\mathcal{H}[s(t)]. \qquad (4.11)$$

For a complex signal $z(t)$, the average angular frequency is calculated as

$$\begin{aligned}
\langle\omega\rangle &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\ \omega|\tilde{z}(\omega)|^2 \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}t\ z^*(t)\int_{-\infty}^{\infty} \mathrm{d}t'\ z(t')\int_{-\infty}^{\infty} \mathrm{d}\omega\ \omega e^{i(t'-t)\omega} \\
&= \frac{1}{2\pi i}\int_{-\infty}^{\infty} \mathrm{d}t\ z^*(t)\int_{-\infty}^{\infty} \mathrm{d}t'\ z(t')\int_{-\infty}^{\infty} \mathrm{d}\omega\ \frac{\partial}{\partial t'}e^{i(t'-t)\omega} \\
&= \frac{1}{i}\int_{-\infty}^{\infty} \mathrm{d}t\ z^*(t)\int_{-\infty}^{\infty} \mathrm{d}t'\ z(t')\frac{\partial}{\partial t'}\delta(t-t') \\
&= \frac{1}{i}\int_{-\infty}^{\infty} \mathrm{d}t\ z^*(t)\dot{z}(t). \qquad (4.12)
\end{aligned}$$

Expressing it in polar form yields

$$\begin{aligned}
\langle\omega\rangle &= \frac{1}{i}\int_{-\infty}^{\infty} \mathrm{d}t\ a(t)e^{-i\phi(t)}\left(\dot{a}(t)e^{i\phi(t)} + ia(t)\dot{\phi}(t)e^{i\phi(t)}\right) \\
&= \int_{-\infty}^{\infty} \mathrm{d}t\ \left(\dot{\phi}(t) - i\frac{\dot{a}(t)}{a(t)}\right)a(t)^2. \qquad (4.13)
\end{aligned}$$

The second term must be zero for $\langle\omega\rangle$ to be real. As a result, $\langle f\rangle$ can be written as

$$\begin{aligned}\langle f\rangle &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}t\ \dot{\phi}(t)|z(t)|^2 \\ &= \frac{1}{2\pi}\left\langle \frac{\mathrm{d}\phi(t)}{\mathrm{d}t}\right\rangle .\end{aligned} \tag{4.14}$$

Therefore, defining instantaneous frequency (IF) by

$$\mathrm{IF}(t) = \frac{1}{2\pi}\frac{\mathrm{d}\phi(t)}{\mathrm{d}t} \tag{4.15}$$

is a natural choice.

For practical demonstration, let us calculate the instantaneous frequency of the sum of two sinusoids,

$$z(t) = a_1 e^{-i\omega_1 t} + a_2 e^{-i\omega_2 t}. \tag{4.16}$$

The Fourier transform of $z(t)$ results in

$$\tilde{z}(\omega) = 2\pi(a_1\delta(\omega-\omega_1) + a_2\delta(\omega-\omega_2)). \tag{4.17}$$

The instantaneous frequency is then calculated as

$$f(t) = -\frac{1}{4\pi}\left\{(\omega_1+\omega_2) - (\omega_1-\omega_2)\frac{a_2^2-a_1^2}{a(t)^2}\right\}, \tag{4.18}$$

where

$$a(t) = \sqrt{a_1^2 + a_2^2 + 2a_1a_2\cos(\omega_1-\omega_2)t}. \tag{4.19}$$

Figure 4.1 plots the instantaneous frequency in Eq. (4.18) for $a_1 = 1$, $a_2 = -2$, $\omega_1 = 2\pi\times 3$, and $\omega_2 = 2\pi\times 5$. Although the individual sinusoids have constant frequencies, the calculated instantaneous frequency varies with time. Furthermore, despite the spectrum of the analytic signal being zero for negative frequencies, the calculated instantaneous frequency can be negative. These observations imply that the instantaneous frequency defined by Eq. (4.15) does not always satisfy our intuition.

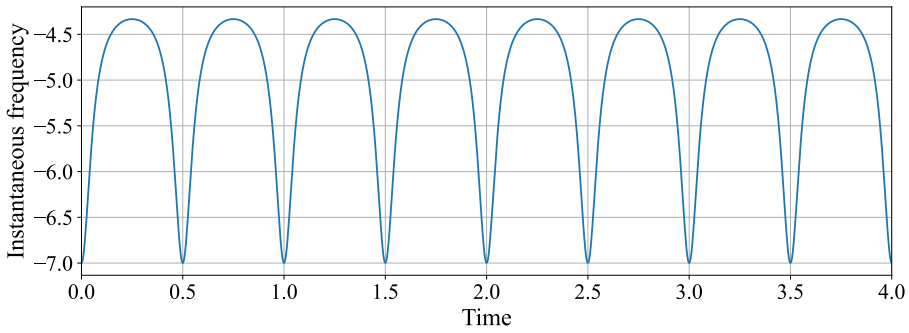


FIG. 4.1. Instantaneous frequency of the sum of two sinusoids described by Eq. (4.16) for $a_1 = 1$, $a_2 = -2$, $\omega_1 = 2\pi\times 3$, and $\omega_2 = 2\pi\times 5$.

4.2 Empirical mode decomposition

We encountered an example in which the instantaneous frequency defined by Eq. (4.15) does not have a physically meaningful component. In response, Huang *et al.* [56] empirically established conditions under which meaningful instantaneous frequencies can be extracted from signals. These conditions are:

(i) The number of extrema and the number of zero crossings must either equal or differ by, at most, one.

(ii) The mean values of the upper and lower envelopes, determined using local maxima and minima, must be zero across all data points.

A function meeting these conditions is termed the *intrinsic mode function*. The process of decomposing a signal into intrinsic mode functions is referred to as the *sifting process*. The procedure to extract the ith intrinsic mode function, denoted as $c_i(t)$, from a signal $x(t)$ is as follows:

1. Let $g(t) = x(t) - \sum_{j<i} c_j(t)$.
2. Identify the maxima and minima of $g(t)$.
3. Generate the upper envelope $u(t)$ and the lower envelope $l(t)$ of $g(t)$ using the cubic spline interpolation (see Appendix E).
4. Calculate the mean values of $u(t)$ and $l(t)$:
$$m(t) = \frac{u(t) + l(t)}{2}. \tag{4.20}$$
5. Subtract $m(t)$ from $g(t)$.
6. Verify if $g(t)$ meets the convergence criterion. We adopt the stoppage criterion based on the Cauchy-convergence test employing a predetermined value ϵ [56, 173]:
$$\frac{\sum_j |m(t_j)|^2}{\sum_j |g(t_j)|^2} < \epsilon. \tag{4.21}$$
If the stoppage criterion is not met, repeat steps 2 to 6 for the updated $g(t)$. Once the conditions are fulfilled, the resulting $g(t)$ is designated as the ith intrinsic mode function.

After several intrinsic mode functions have been extracted, the process terminates if the residue becomes monotonic or exhibits only one extremum. Consequently, we obtain the decomposition
$$x(t) = \sum_i c_i(t) + r(t). \tag{4.22}$$
The sequential steps from 1 through 5 and the resulting intrinsic mode functions for a test signal are plotted in Figs. 4.2 and 4.3, respectively.

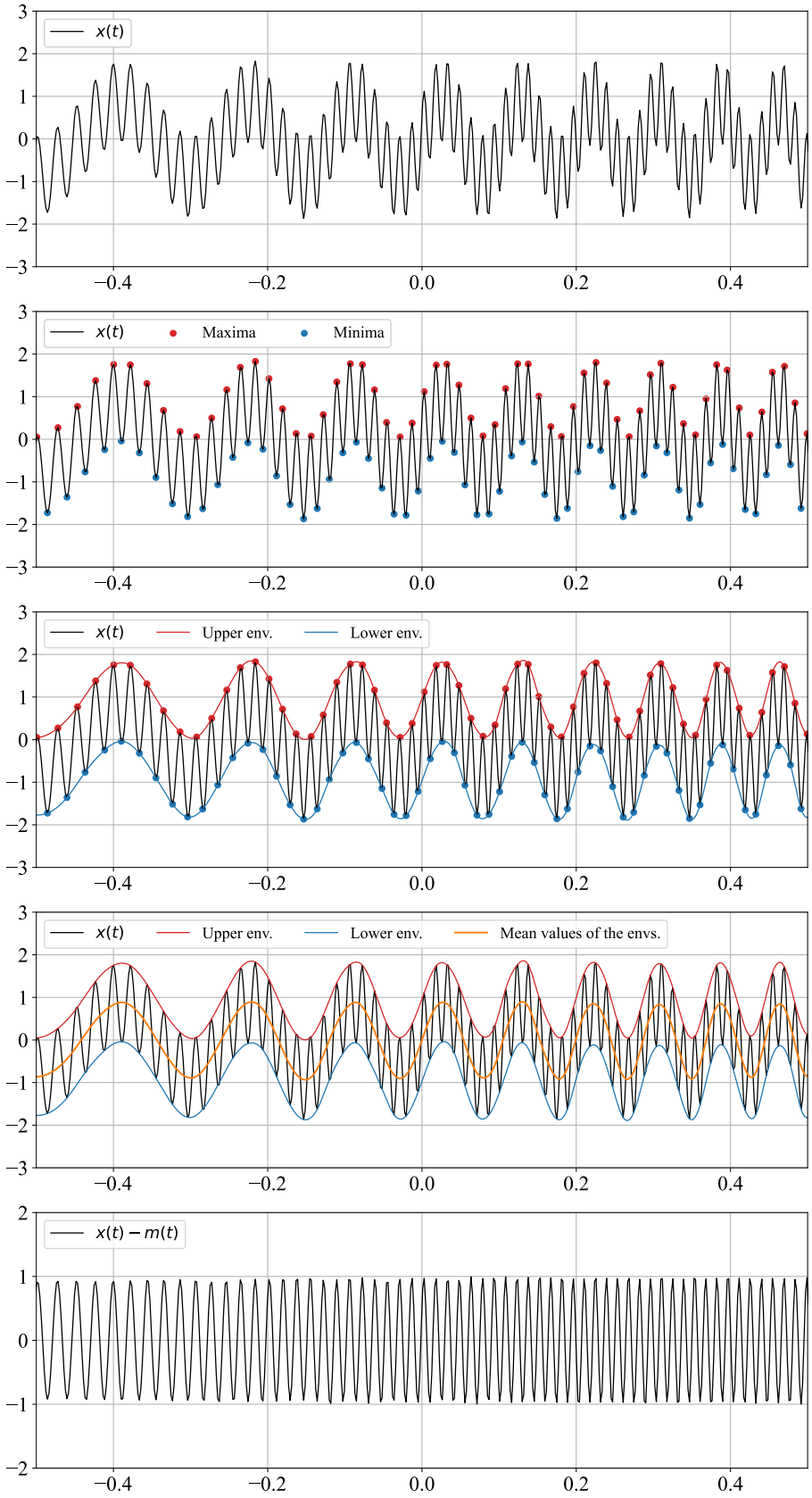


FIG. 4.2. Excerpt of sequential steps in extracting the first intrinsic mode function from a test signal.

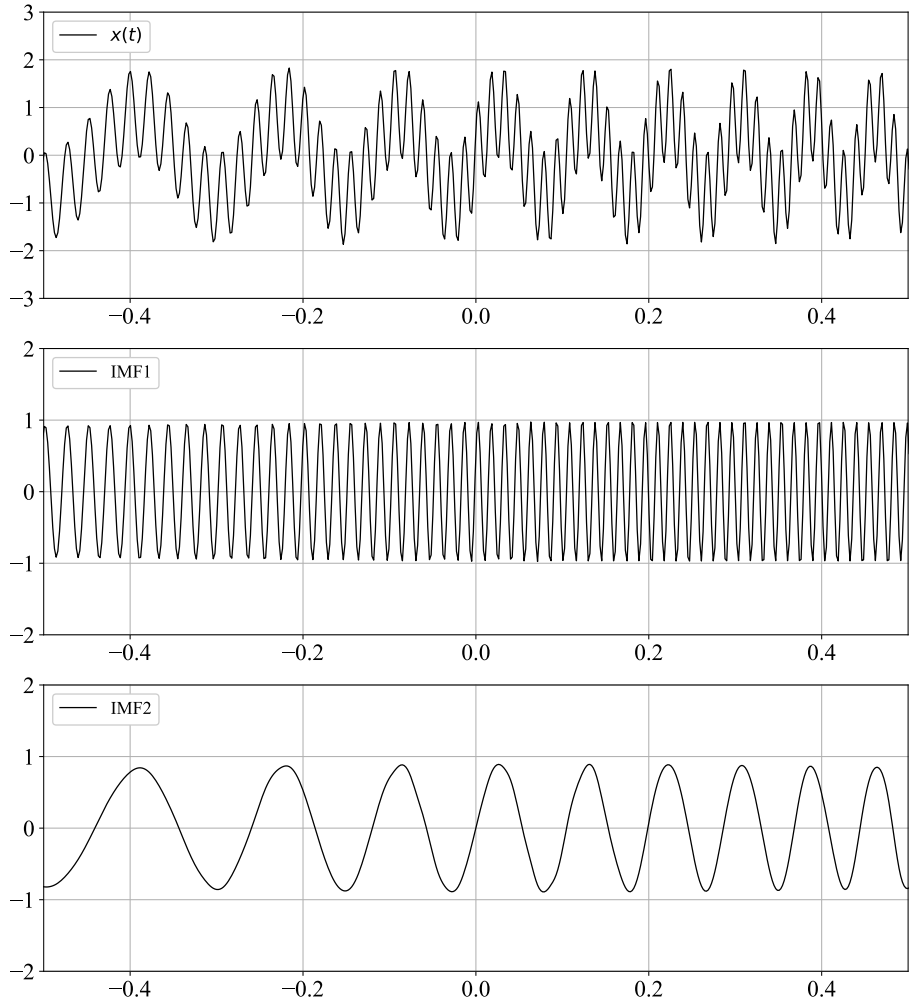


FIG. 4.3. Test signal (top) and the extracted two intrinsic mode functions (middle and bottom) using the sifting process.

4.3 Hilbert spectral analysis

After extracting the intrinsic mode functions using the empirical mode decomposition from a signal, we can further analyze the signal's time-varying spectral characteristics by computing the analytic signal for each intrinsic mode function, along with determining the instantaneous amplitude and the instantaneous frequency. As previously explained, for the ith intrinsic mode function, $c_i(t)$, the instantaneous amplitude $a_i(t)$ and instantaneous frequency $f_i(t)$ are calculated as

$$a_i(t) = \sqrt{c_i(t)^2 + \mathcal{H}[c_i(t)]^2}, \tag{4.23}$$

$$f_i(t) = \frac{1}{2\pi}\frac{\mathrm{d}}{\mathrm{d}t}\tan^{-1}\left(\frac{\mathcal{H}[c_i(t)]}{c_i(t)}\right). \tag{4.24}$$

Since the instantaneous amplitude and instantaneous frequency of each intrinsic mode function are obtained as functions of time, it is possible to visualize these information in a three-dimensional space of time, frequency, and amplitude. In particular, one can

construct a time-frequency map by

$$H(t, f) = \sum_i a_i(t)\delta(f - f_i(t)). \tag{4.25}$$

This representation is termed the *Hilbert spectrum*, and the methodology employed to generate it is known as the *Hilbert spectral analysis*. The process of deriving the Hilbert spectrum from a signal through empirical mode decomposition and Hilbert spectral analysis is referred to as the Hilbert-Huang transform.

Figure 4.4 shows the instantaneous amplitude and the instantaneous frequency of each intrinsic mode function of the test signal used in the previous section. Figure 4.5 exhibits the time-frequency maps of the test signal computed using both the short-time Fourier transform (STFT) and the Hilbert-Huang transform. The time and frequency resolutions of the STFT map are 0.01 and 0.5, respectively. These resolutions are constrained by the uncertainty relationship. In contrast, the Hilbert-Huang transform computes the instantaneous amplitude and instantaneous frequency for each data point, offering superior resolution compared to the STFT.

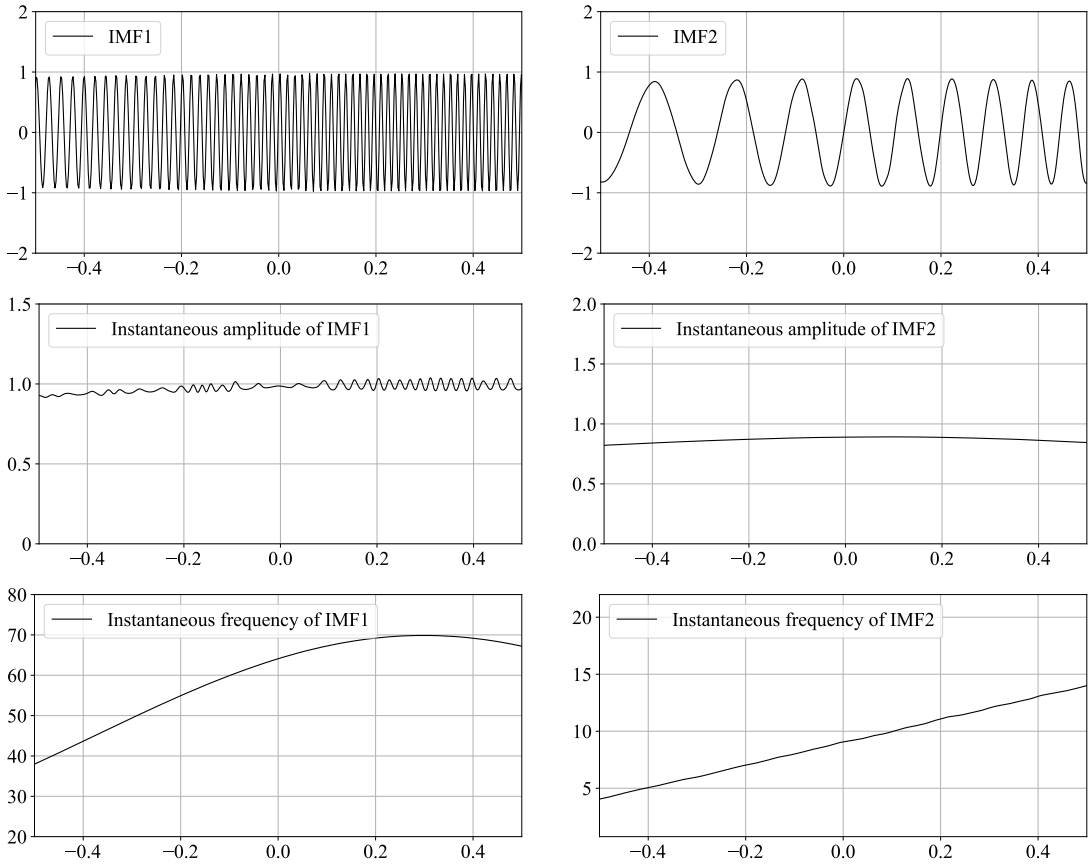
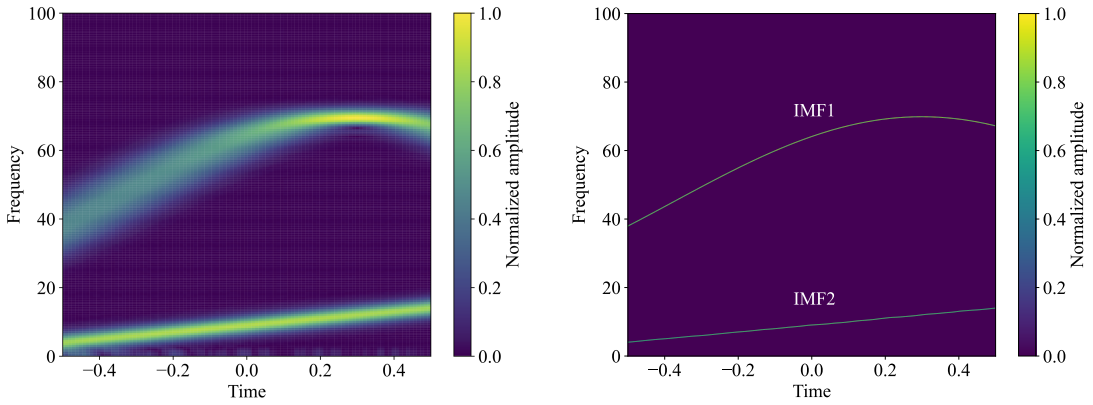


FIG. 4.4. Instantaneous amplitude and instantaneous frequency of the intrinsic mode functions of the test signal.



(a) Short-time Fourier transform (b) Hilbert-Huang transform

FIG. 4.5. Time-frequency maps of the test signal generated using the short-time Fourier transform and the Hilbert-Huang transform.

4.4 Advanced techniques

4.4.1 Ensemble empirical mode decomposition

Despite its effectiveness in decomposing signals into intrinsic mode functions, the empirical mode decomposition faces several challenges that can limit its accuracy and reliability in certain situations. One major drawback is *mode mixing*, wherein different intrinsic modes get combined, making it challenging to extract individual meaningful components accurately. To address this challenge, a noise-assisted method, known as the *ensemble empirical mode decomposition*, was introduced as an enhancement to provide more robust signal decomposition [174]. The method involves enhancing a signal by incorporating white Gaussian noise to create multiple samples and derive the intrinsic mode functions from each sample. The subsequent step is to compute the average of these intrinsic mode functions across different noise-added samples. The algorithmic steps for the ensemble empirical mode decomposition are as follows:

1. Generate N sequences, denoted as $n_k(t)$ for $k = 1, \ldots, N$, where each sequence represents a Gaussian noise with a mean of zero and a predetermined standard deviation σ.

2. Construct an ensemble of noise-added data by adding the noise sequences to the target data $s(t)$. Formally, the kth data is represented as

$$s_k(t) = s(t) + n_k(t). \tag{4.26}$$

3. Apply the empirical mode decomposition to each $s_k(t)$ to obtain the intrinsic mode functions, along with the residue $r_k(t)$,

$$s_k(t) = \sum_i c_{i,k}(t) + r_k(t). \tag{4.27}$$

Here, $c_{i,k}(t)$ is the ith intrinsic mode function of $s_k(t)$.

4. Determine the definitive intrinsic mode functions and residue by taking the ensemble mean,

$$c_i(t) = \frac{1}{N}\sum_{k=1}^{N} c_{i,k}(t), \quad r(t) = \frac{1}{N}\sum_{k=1}^{N} r_k(t). \tag{4.28}$$

The process of the Hilbert-Huang transform with the ensemble empirical mode decomposition is illustrated in Fig. 4.6.

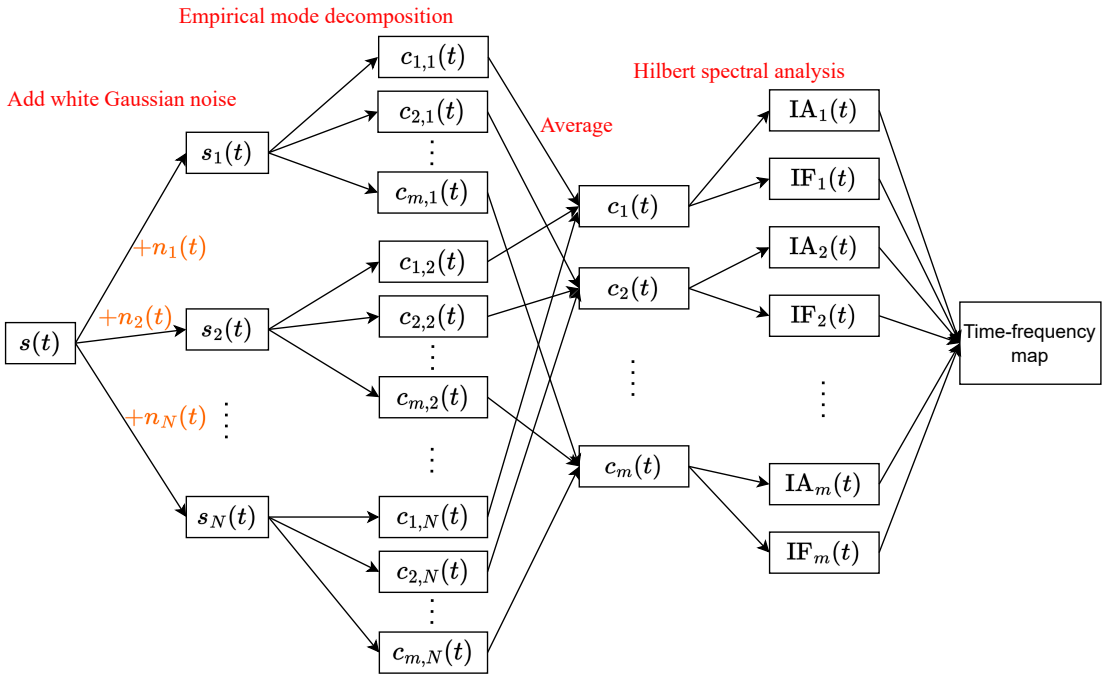


FIG. 4.6. Schematic diagram of the ensemble empirical mode decomposition. IA_k and IF_k denote the instantaneous amplitude and the instantaneous frequency of kth intrinsic mode function $c_k(t)$, respectively.

4.4.2 Complementary ensemble empirical mode decomposition

In the ensemble empirical mode decomposition, the residue resulting from the addition of white noise, defined as the difference between the original data and the reconstructed data, is proportional to $1/\sqrt{N}$, where N represents the number of trials used to derive the ensemble intrinsic mode functions. An alternative approach was proposed by Yeh *et al.* [175], wherein white noise is added in pairs—both positive and negative—to the original data, thereby generating two sets of ensemble intrinsic mode functions. The final intrinsic mode function is the ensemble of both the intrinsic mode functions with positive and negative noises. This method reduce the computational costs and the residue originated from white noises. The algorithmic steps of this technique is as follows:

1. Generate N sequences, denoted as $n_k(t)$ for $k = 1, \ldots, N$, where each sequence represents a Gaussian noise with a mean of zero and a predetermined standard deviation σ.

2. Construct an ensemble of noise-added data $s_k^{(+)}(t)$ and noise-subtracted data $s_k^{(-)}(t)$ to the target data $s(t)$ for $k = 1, \ldots, N$.

$$s_k^{(+)}(t) = s(t) + n_k(t), \tag{4.29}$$

$$s_k^{(-)}(t) = s(t) - n_k(t). \tag{4.30}$$

3. Apply the empirical mode decomposition to each $s_k^{(+)}(t)$ and $s_k^{(-)}(t)$ to obtain the intrinsic mode functions, along with the residue $r_k^{(\pm)}(t)$,

$$s_k^{(+)}(t) = \sum_i c_{i,k}^{(+)}(t) + r_k^{(+)}(t), \tag{4.31}$$

$$s_k^{(-)}(t) = \sum_i c_{i,k}^{(-)}(t) + r_k^{(-)}(t). \tag{4.32}$$

   Here, $c_{i,k}^{(\pm)}(t)$ is the ith intrinsic mode function of $s_k^{(\pm)}(t)$.

4. Determine the definitive intrinsic mode functions and residue by

$$c_i(t) = \frac{1}{2N} \sum_{k=1}^{N} \left(c_{i,k}^{(+)}(t) + c_{i,k}^{(-)}(t) \right), \quad r(t) = \frac{1}{2N} \sum_{k=1}^{N} \left(r_k^{(+)}(t) + r_k^{(-)}(t) \right). \tag{4.33}$$

4.5 Hilbert-Huang transform in gravitational-wave astronomy

Unlike the STFT or the wavelet transform, which are susceptible to time-frequency uncertainty, the Hilbert-Huang transform is able to generate a high-resolution time-frequency representation. This characteristic makes it particularly effective for analyzing non-stationary and nonlinear signals, making it a valuable tool for GW analysis without assumptions on a waveform model. In addition to the application to signal detection, its ability to compute an instantaneous frequency with a function of time makes it a valuable tool for studying underlying physics behind the GW waveforms.

The first application of the Hilbert-Huang transform in GW data analysis was performed by Camp *et al.* [176]. They computed the instantaneous frequency of the inspiral of a supermassive black hole merger and demonstrated the effectiveness of the Hilbert-Huang transform to burst GW analysis. Detection method of BBH mergers using the Hilbert-Huang transform was developed in Ref. [177]. They incorporated multiple techniques such as Bayesian blocking and kernel density estimation to detect and characterize low-SNR signals. In the subsequent study, Stroeer *et al.* [178] used instantaneous frequency and amplitude to determine the relative timing delay of signals in multiple detectors. Takahashi *et al.* [179] investigated the optimal parameters for the ensemble empirical mode decomposition and method for extracting local extrema by using sine-Gaussian test signals. It was concluded that the stoppage criterion highly affects the accuracy of extracted instantaneous frequency.

Since the first detection of GW events, several studies explored the application the Hilbert-Huang transform to real detected data. Sakai *et al.* [170] successfully extracted

the GW150914 signal from observed data by applying a bandpass filter and a bandstop filter before the ensemble empirical mode decomposition. Various observed signals in O1 through O3 were analyzed in Ref. [180]. Among the analyzed signals, they reported that they could not extract the two BNS signals, GW170817 and GW190425, due to their small amplitudes. These results are consistent with another study in Ref. [181].

Applications to analyze post-merger signals of CBC events have been explored in several studies. Kaneyama *et al.* [182] analyzed post-merger BNS signals and demonstrated the classification of equation of states by accurately estimating the frequency. This method was improved by Yoda *et al.* [173] by using Akima spline interpolation [183] instead of the conventional cubic spline interpolation in the empirical mode decomposition. Estimating starting times and frequencies of quasinormal modes from BBH ringdown was performed in Ref. [171]. The method was compared to other approaches such as neural networks and autoregressive models in Ref. [184].

For analyzing GW signals from CCSNe, Takeda *et al.* [57] used the Hilbert-Huang transform to extract GWs associated with the SASI. They estimated the starting time and the frequencies of the SASI mode from a simulated signal of neutrino-driven supernova explosion. Hu *et al.* [181] also analyzed neutrino-driven CCSN signals and indicated the existence of the SASI mode in the time-frequency map. Yuan *et al.* [185] used the ensemble empirical mode decomposition to both neutrino-driven and magnetorotational CCSN signals to reconstruct waveforms.

Part II

Research Results

Chapter 5

Visualizing Convolutional Neural Network for Classifying Gravitational Waves from Core-Collapse Supernovae

This chapter is a reprint of the paper [2] with minor modifications in formatting. Figure 5.2 has been regenerated to match the dimensions of this thesis.

Abstract

In this study, we employ a convolutional neural network to classify gravitational waves originating from core-collapse supernovae. Training is conducted using spectrograms derived from three-dimensional numerical simulations of waveforms, which are injected onto real noise data from the O3 observation of both Advanced LIGO and Advanced Virgo. To gain insights into the decision-making process of the model, we apply class activation mapping techniques to visualize the regions in the input image that are significant for the model's prediction. The class activation maps reveal that the model's predictions predominantly rely on specific features within the input spectrograms, namely the g-mode and low-frequency modes. The visualization of convolutional neural network models provides interpretability to enhance their reliability and offers guidance for improving detection efficiency.

5.1 Introduction

The first detection of gravitational waves (GWs) from a bianry black hole merger by the Advanced Laser Interferometer Gravitational-wave Observatory (Advanced LIGO) [5] in 2015 marked the beginning of GW astronomy [6]. Throughout three observing runs (O1, O2, and O3), Advanced LIGO and Advanced Virgo [18] reported 90 GW events [31–34]. As of May 2023, the international GW network, now including KAGRA [19], has begun its fourth observing run (O4) with improved sensitivity.

All of the GW events detected so far are exclusively from compact binary coalescences. However, short-duration GW bursts arising from core-collapse supernovae (CCSNe) are

expected to be detected by the current and next-generation GW detectors, such as the Einstein Telescope [66] and the Cosmic Explorer [68]. CCSNe, resulting from massive star explosions leading to neutron stars or stellar-mass black holes, stand as one of the most energetic astrophysical events in the Universe, emitting electromagnetic waves, neutrinos, and GWs. While electromagnetic waves from CCSNe are frequently observed, neutrinos have only been detected from SN1987A [38, 39]. GWs are expected to carry information about the inner core's dynamics, providing vital insights into the explosion mechanism which remains elusive. The primary conundrum lies in discerning how a stalled shock wave is revived to cause a star to explode. Currently, there are two prevailing theories [91]: the neutrino-driven mechanism [186], in which shock waves are revived by neutrinos stored behind the shock wave heating the surrounding matter, and the magnetorotational mechanism [187], in which the rapid rotation of the progenitor causes explosions driven by strong magnetic fields. The typical GW detection range for neutrino-driven signals is expected to be around 10 kpc, while the detection range for magnetorotational signals is expected to be above 100 kpc [188].

Due to the stochastic nature of GW signals from CCSNe, the conventional matched-filtering technique, which relies on specific waveform templates, is unsuitable. Alternative detection methods based on the time-frequency representation have been devised in response. In particular, the coherent WaveBurst pipeline [148, 149] detects and reconstructs burst GW signals by searching for excess power in a time-frequency map, with minimal reliance on a specific source model.

Predicting GW signals from CCSNe remains a formidable challenge. However, recent advancements in theoretical research and multidimensional numerical simulations have revealed certain signal properties. For neutrino-driven CCSNe, the dominant emissions arise from the g-mode oscillation of the proto-neutron star surface. These frequencies progressively increase over time, ranging from a few hundred Hz to a few kHz. Additionally, at low frequencies ($\lesssim$ 200 Hz), GW emissions associated with hydrodynamics instabilities including neutrino-driven convection and standing accretion shock instability (SASI) [94] are observed in some simulations. Insights obtained from these simulations are pivotal in enhancing methods for CCSNe detection and analysis.

In recent years, machine learning techniques, especially deep learning, have gained traction in a variety of scientific fields due to their capacity for recognizing intricate patterns and extracting meaningful features from large data sets. This ability has been especially noted in areas such as computer vision and natural language processing. Its application in GW research has followed, with numerous implementations and explorations, as highlighted in a comprehensive review in Ref. [41] and the foundational efforts by George and Huerta [44, 189]. In the field of CCSNe analysis, Astone *et al.* [46] leveraged convolutional neural networks (CNNs) to detect CCSNe within Gaussian noise, using g-mode phenomenological waveforms for training, outperforming the coherent WaveBurst pipeline. Consecutive studies by Iess *et al.* [150, 152] involved the training of both one- and two-dimensional CNNs and long short-term memory networks [190] to identify seven distinct CCSN waveforms embedded in real noise and glitches, with their models achieving 98% classification accuracy at 1 kpc with a three-detector network. Additionally, Chan *et al.* [151] employed one-dimensional CNNs to investigate both magnetorotational and

neutrino-driven signals in Gaussian noise, recording a true alarm probability of 80% for magnetorotational signals from sources at 60 kpc and 55% for neutrino-driven signals from sources at 10 kpc with a fixed false alarm probability of 10%. In another study, Edwards [153] used two-dimensional CNNs to classify 18 different equations of state from pure magnetorotational CCSN signals, attaining an accuracy of 72%. López *et al.* [47] refined phenomenological waveforms originally used by Astone *et al.* [46], achieving a 60% true alarm probability for signals located at 15 kpc with a 5% false alarm rate.

Although deep learning exhibits strong performance on a wide range of tasks, its intricate models, characterized by a large number of parameters, pose challenges in elucidating their decision-making processes. To address this, the field of explainable artificial intelligence [55] has surged, aiming to make model decisions transparent and interpretable. Within the context of CNNs, efforts have been made to develop techniques that attempt to understand the decision-making process by reverse mapping the output of the network into the input space to identify the specific input components that were discriminative in producing the output. Class activation mapping (CAM) [129] is one such method, which computes a weighted sum of the outputs of the last convolutional layer using the outputs of the global average pooling layer after the last convolutional layer as weights. It helps identify the regions in the input image that were important for a prediction, but the model needs to be modified to include a global average pooling layer, which may result in lower accuracy. Grad-CAM [130] was introduced as a solution to this limitation of CAM, offering the advantage of not requiring any modifications to the network architecture by using gradient information from the prediction for weighted parameters. Subsequently, Grad-CAM++ [131], a generalization of Grad-CAM, and Score-CAM [132], a gradient-free CAM method, were developed to generate more accurate saliency maps than Grad-CAM. These techniques to analyze deep learning models are commonly used in fields such as electrocardiogram signal analysis [191] and X-ray diagnosis [192]; however, for GW analysis, they have only been used in Ref. [193] to the best of our knowledge.

In this study, we first take an approach similar to Ref. [152] and train a two-dimensional CNN model to classify CCSNe signals using short-time Fourier-transformed spectrograms as input for simplicity. We use nine types of waveforms from recent three-dimensional numerical simulations and O3 real noise to train and validate our model. In the test, signals from sources between 1 and 10 kpc are considered, and the performance of the model for sources at each distance is discussed. To interpret the model, we use three CAM methods to generate saliency maps and evaluate them using two metrics: average drop and average increase. The best CAM method is then applied to correctly classified and also misclassified samples to visualize the regions in the input spectrogram that influence the predictions of our model.

The remainder of this chapter is organized as follows. Section 5.2 describes our data sets, the CNN model, and the CAM techniques. In Sec. 5.3, we discuss the classification performance of our model, and apply multiple visualization techniques to interpret the model. We summarize and conclude this chapter in Sec. 5.4.

5.2 Method

Our CNN model is trained to classify strains at the three detectors LIGO Hanford (H1), LIGO Livingston (L1), and Virgo (V1) into ten classes: noise and nine different CCSN waveforms. In this section, we first provide an overview of the data used in this study, including a brief summary of the CCSN simulation data and the preprocessing strategy to generate our training, validation, and test sets. Subsequently, our CNN architecture and the theory of the visualization technique of the model are explained.

5.2.1 Data set

CCSN waveforms

Modeling the stellar core collapse, bounce, and subsequent post-bounce evolution is very complicated and computationally expensive. However, remarkable advancements in three-dimensional numerical simulations of neutrino-driven explosions have been achieved by several groups in recent years. Specific details of the waveform depend on various properties of the progenitor, such as the mass, angular velocity, and equation of state of the dense matter. Both the general relativity approximation and the handling of neutrino transport critically influence simulations. From the available simulation data under a variety of conditions, we select nine types of waveforms from four recent three-dimensional numerical simulations [194–197]. All of them allow us to compute the GW amplitude in any observer direction from the quadrupole moment.

Powell and Müller [194] performed simulations using the general-relativistic neutrino hydrodynamics code CoCoNuT-FMT [198]. We use two waveforms from the models `he3.5` and `s18`. The progenitor of `he3.5` is an ultra-stripped star evolved from a helium star with an initial mass of 3.5 $M_\odot$. The simulation is stopped at 0.7 s after core bounce. The GW is dominated by excitation of g-modes in the proto-neutron star with a peak frequency around 900 Hz. Model `s18` is a single star with a zero-age main-sequence mass of 18 $M_\odot$. The simulation was stopped 0.89 s after core bounce. The GW emission is similar to model `he3.5`, with g-modes oscillations of the proto-neutron star with a peak frequency around 900 Hz.

Radice *et al.* [195] studied eight models using the Eulerian radiation-hydrodynamics code FORNAX [199]. We use waveforms from the models `s13` and `s25` corresponding to progenitors of 13 and 25 $M_\odot$ zero-age main-sequence, respectively. The simulation is ended at 0.77 s in `s13` and 0.62 s in `s25` after bounce. Both waveforms are characterized by f- and g- modes with a peak frequency around 1400 Hz in `s13` and 1100 Hz in `s25`. In addition, the `s25` waveform has a clear SASI mode around 100 Hz.

From the simulation by Powell and Müller [196], we use the three models `m39`, `y20`, and `s18np`. Model `m39` is a rapidly rotating 39 $M_\odot$ Wolf-Rayet star with an initial surface rotation velocity of 600 km s^{-1}. It produces a neutrino-driven explosion without magnetic fields. The other two are nonrotating models of 20 $M_\odot$ Wolf-Rayet star and an 18 $M_\odot$ zero-age main-sequence star. The simulation is ended at 0.98, 1.2, 0.56 s after core bounce in models `m39`, `y20`, and `s18np`, respectively. All three models show GW emission associated with prompt convection shortly after bounce as well as f-mode oscillations of the proto-neutron star. In model `s18np`, the absence of strong perturbations from convective

oxygen burning, in contrast to `s18`, prevents the shock from being revived and leads to the development of strong SASI activity, with a frequency reaching $\sim$ 400 Hz by the end of the simulation.

Powell *et al.* [197] performed simulations using three equation of states: LS220 [200], SFHx, and SFHo [201]. The progenitor models are 85 and 100 $M_\odot$ Population III zero-age main-sequence stars. We use the `z85_sfhx` and `z100_sfho` models. The simulation is ended at 0.59 s in `z85_sfhx` and 0.62 s in `z100_sfho`. Both waveforms show typical g-mode emission with a peak frequency of $\sim$ 700 Hz. In the `z85_sfhx` model, the frequency of the SASI emission increases up to the point of shock revival, reaching $\sim$ 200 Hz, and decreases afterwards. In the `z100_sfho` model, which does not explode, the frequency of the SASI emission continues to increase by the end of the simulation, reaching $\sim$ 400 Hz.

Figure 5.1 shows the amplitude spectral density of the plus mode of each waveform at 1 kpc from the polar direction. The amplitude of the `m39` waveform is the largest of these waveforms. Waveforms that have peaks around 100 Hz such as `s25` and `s13` indicate that they have SASI-induced GW modes.

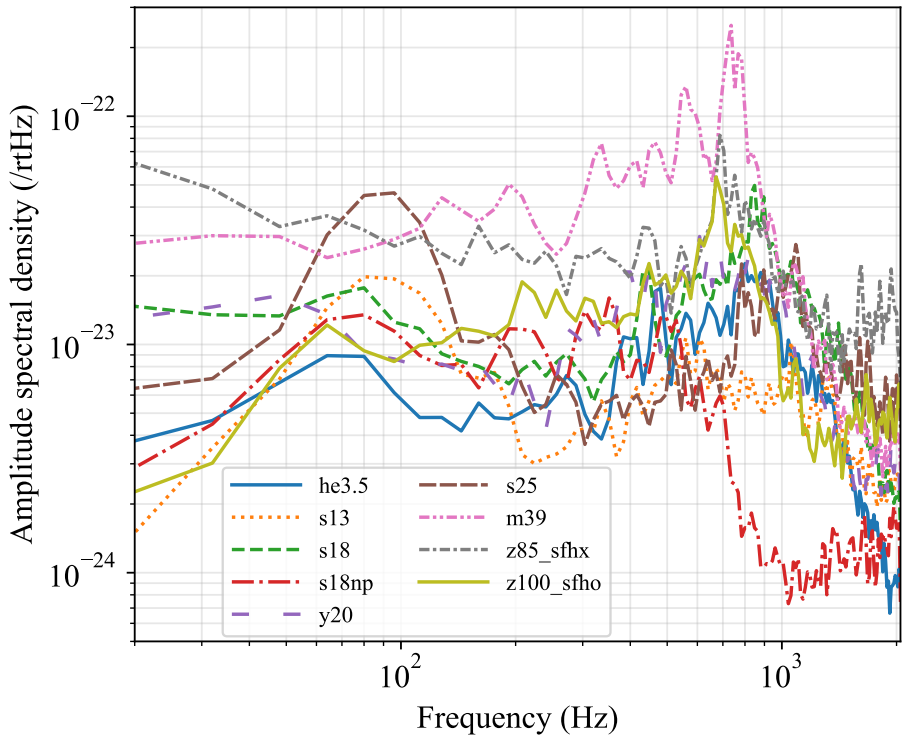


FIG. 5.1. Amplitude spectral density of the plus mode of each waveform at 1 kpc. The observer is in the polar direction.

Data processing

Our data sets are generated from the simulation data outlined in the previous section. A crucial step in this process is the computation of the GW amplitude. This is accomplished by uniformly sampling the direction of radiation (θ, ϕ) and using the formulae for

the plus and cross polarization

$$h_+ = \frac{1}{D}\frac{G}{c^4}(\ddot{Q}_{\theta\theta} - \ddot{Q}_{\phi\phi}), \tag{5.1}$$

$$h_\times = \frac{1}{D}\frac{2G}{c^4}\ddot{Q}_{\theta\phi}, \tag{5.2}$$

where Q is the traceless quadrupole moment and D is the distance between a source and Earth. As the sampling of the simulation is usually not uniform in time, we resample data uniformly with a sampling rate of 4096 Hz. A high-pass filter with a cutoff frequency of 11 Hz and a Tukey window with $\alpha = 0.1$ are applied to the resampled signals. Each signal is then truncated or padded with zeros to make the length 1 second. In order to make the model robust, we randomly time shift the signals so that the time of the core bounce is between 0 and 0.15 s. For the training and validation sets, the signals are scaled using an optimal matched filter signal-to-noise ratio (SNR), as the amplitudes of the simulated signals are quite different. The SNR is defined as

$$\rho = \sqrt{4\int_{f_{\min}}^{f_{\max}} \mathrm{d}f \ \frac{|\tilde{h}(f)|^2}{S_n(f)}}, \tag{5.3}$$

where $\tilde{h}(f)$ is the Fourier transform of the signal and $S_n(f)$ is the one-sided power spectral density of the noise. The network SNR of the detectors H1, L1, and V1, given by

$$\rho_{\mathrm{net}} = \sqrt{\rho_{\mathrm{H1}}^2 + \rho_{\mathrm{L1}}^2 + \rho_{\mathrm{V1}}^2}, \tag{5.4}$$

is used to scale the signals. We generate samples with network SNRs from 20 to 50 for training and validation sets. For the test set, the signals are scaled to have distances between 1 and 10 kpc. Sky location is also randomly selected and the GW amplitude $h(t)$ is computed, taking into account the antenna pattern functions F_+ and $F_\times$ and the delay in arrival time of each detector with the following equation:

$$h(t) = F_+(\alpha, \delta, \psi, t)h_+(t + \Delta t) + F_\times(\alpha, \delta, \psi, t)h_\times(t + \Delta t), \tag{5.5}$$

where α is the right ascension, δ is the declination, and ψ is the polarization angle. Δt is the delay in arrival time between the detector and the center of the Earth. We use the PyCBC software library [202] to carry out these computations.

Noise used in this study is O3 real data of Advanced LIGO and Advanced Virgo, obtained from the Gravitational Wave Open Science Center [203]. Data from GPS time 1238236470 to 1238252308 is used for the training set, 1238265720 to 1238354855 is used for the validation set, and 1238404064 to 1238457121 is used for the test set. Data around the event time reported in the second Gravitational Wave Transient Catalog [32] are excluded. After a signal is injected into the noise, each sample is whitened with the power spectral density computed using Welch's method [204] and then short-time Fourier transformed with a window size of 0.0625 seconds to produce a spectrogram. The spectrogram is normalized to [0, 1] before input to the network.

We generated 60 000 samples for the training and validation sets, and 100 000 samples for the test set. The test set has 1000 samples for each class and each distance. Sample

spectrograms in the training set are shown in Fig. 5.2.

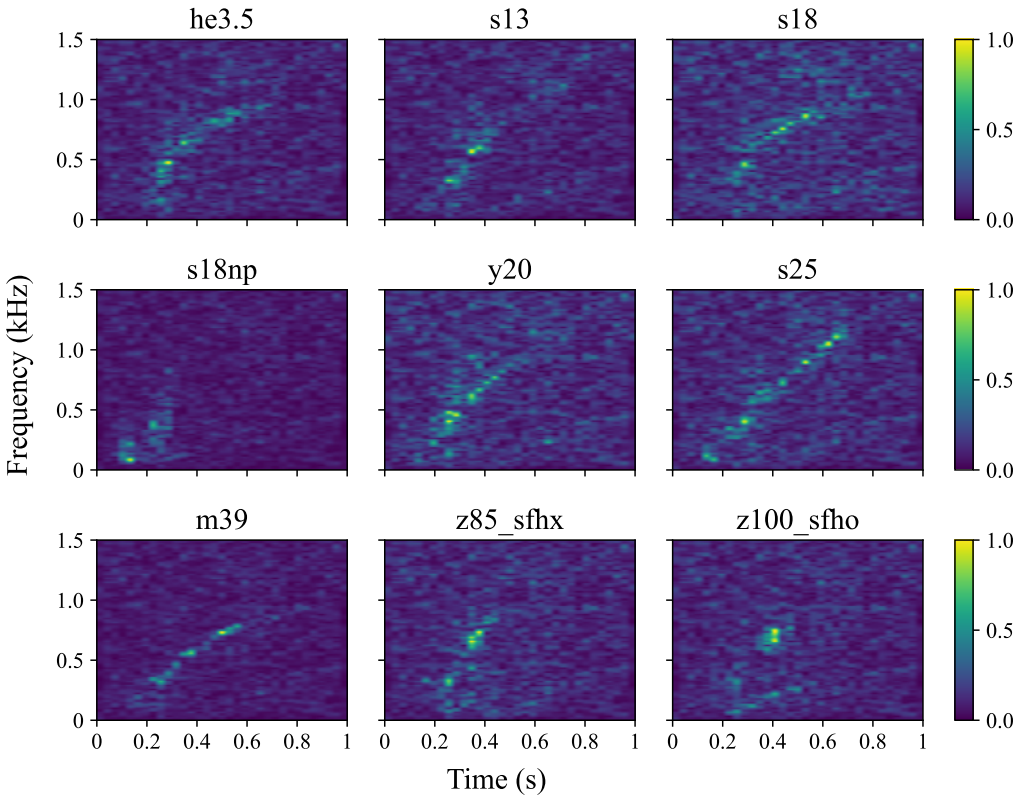


FIG. 5.2. Sample whitened spectrograms of each class at the H1 detector in the training set. Each signal sample is observed in the polar direction and scaled to have an SNR of 40. The bounce time is fixed at 0.1 s.

5.2.2 CNN model

Our CNN model consists of two convolutional layers of kernel size 3, each followed by a max-pooling layer of size 2 and a ReLU layer. The outputs of these layers are fed into two fully connected layers, and finally the softmax layer outputs a size-10 vector whose elements represent a probability of each class. The model has 427 378 trainable parameters in total. This model is shallower than the one used in Ref. [152]. However, its classification performance is comparable to the previous study, prompting us to adopt this model. Reducing the number of layers also helps us generate higher-resolution CAM maps.

The model is trained using categorical cross entropy as the loss function and Adam optimizer [123] with a learning rate of 5×10^{-4} to update the weights. In the training, we adopt curriculum learning [205] as a strategy to enhance the model and accelerate the training by starting from inputting high-SNR samples and gradually adding lower-SNR samples. We train the model on a single GPU (NVIDIA GeForce RTX3090) for 120 epochs with a mini-batch size of 128.

5.2.3 Visualization

After training the model, we use CAM techniques to generate saliency maps. These maps show the regions in the input that influenced the model's prediction. In this study, we select three CAM methods—Grad-CAM, Grad-CAM++, and Score-CAM—which are widely used today to interpret CNN models. All of these CAM techniques are applied to the convolutional layer prior to the final max-pooling layer in our model.

Grad-CAM

Grad-CAM is a gradient-based visualization technique that highlights the important regions of an input image that the model is looking at while making a prediction. Suppose that for a given input, the prediction score for class c before the softmax layer of the trained model is y^c, and the kth output matrix of the last convolutional layer is A^k. To obtain the Grad-CAM map of class c, we first compute the gradients of the score y^c with respect to the (i, j) component of the kth feature map A^k. We then take the global average of these gradients:

$$\alpha_k^c = \frac{1}{Z} \sum_{i,j} \frac{\partial y^c}{\partial A_{ij}^k}, \tag{5.6}$$

where Z is the number of pixels in A^k. This weight α_k^c represents the importance of the feature map k for the class c.

The Grad-CAM map of the class c is computed as a linear sum of A^k with α_k^c as weights. The ReLU function is applied to extract only features that have a positive contribution to the prediction score. The resulting map of class c is expressed as

$$L_{\mathrm{Grad-CAM}}^c = \mathrm{ReLU}\left(\sum_k \alpha_k^c A^k \right). \tag{5.7}$$

Since convolutional layers and pooling layers make the size of the feature map smaller than the input, the Grad-CAM map is finally interpolated to make it the same size as the input.

Grad-CAM++

While Grad-CAM takes a global average of the gradient matrix when calculating the weight α_k^c in Eq. (5.6), Chattopadhay *et al.* [131] proposed a method to fully include the importance of each pixel in the gradient matrix by taking its weighted average for the weight:

$$\alpha_k^c = \sum_{i,j} \alpha_{ij}^{kc} \ \mathrm{ReLU}\left(\frac{\partial y^c}{\partial A_{ij}^k} \right). \tag{5.8}$$

The ReLU function is used to account for features that increase the activation of the output neuron rather than suppress the activation of the output neuron. The weights α_{ij}^{kc}

can be theoretically derived using higher-order derivatives:

$$\alpha_{ij}^{kc} = \frac{\frac{\partial^2 y^c}{(\partial A_{ij}^k)^2}}{2\frac{\partial^2 y^c}{(\partial A_{ij}^k)^2} + \sum_{a,b} A_{ab}^k \frac{\partial^3 y^c}{(\partial A_{ij}^k)^3}}. \tag{5.9}$$

This method is known as Grad-CAM++, since it can be considered as a generalization of Grad-CAM. The saliency map for Grad-CAM++ is expressed in the same way as for Grad-CAM, using weights in Eq. (5.8) and feature maps, as

$$L^c_{\mathrm{Grad-CAM++}} = \mathrm{ReLU}\left(\sum_k \alpha_k^c A^k\right). \tag{5.10}$$

Score-CAM

Wang *et al.* [132] proposed a gradient-free CAM method called Score-CAM. It solves the problem of gradient-based CAM methods, namely that the gradient is unstable, easily disturbed by noise, and can vanish or explode in deep networks. To generate a Score-CAM map, feature maps are used to mask an input image. Let H^k be the kth feature map, up-sampled to the same size as the input and normalized to [0, 1]. Given an input image X, the weight for the kth feature map is computed as the difference between the score of the masked image $X \odot H^k$ and the score of the baseline image X_b:

$$\alpha_k = f(X \odot H^k) - f(X_\mathrm{b}), \tag{5.11}$$

where $f(\cdot)$ denotes the output of the CNN and $\odot$ denotes the Hadamard product. A black image is used as a baseline image. The Score-CAM map of class c is then computed as a linear sum of the cth value of α_k and the feature map A^k as

$$L^c_{\mathrm{Score-CAM}} = \mathrm{ReLU}\left(\sum_k \alpha_k^c A^k\right). \tag{5.12}$$

5.3 Results and discussion

5.3.1 Classification performance

Figure 5.3 shows the evolution of the categorical cross entropy loss function during the training. Initially, we input samples with SNRs between 40 and 50. In the subsequent 40 epochs, samples with SNRs between 30 and 40 are input, leading to the temporary increase in the loss at epoch 40. In the last 40 epochs, as we input samples with SNRs between 20 and 30, a similar temporary increase in loss is observed at epoch 80. From the loss curves, we confirm that there is no significant overfitting to the training data.

Classification accuracy is defined as the proportion of correctly classified samples out of the total number of samples. After training, the model achieves a classification accuracy of 97.8% on a validation set consisting of uniformly sampled signals with SNRs between 20 and 50. On the test set, our model shows an accuracy of 98.4% for signals with sources from 1 kpc, which is comparable to the results of the previous study [152], despite some

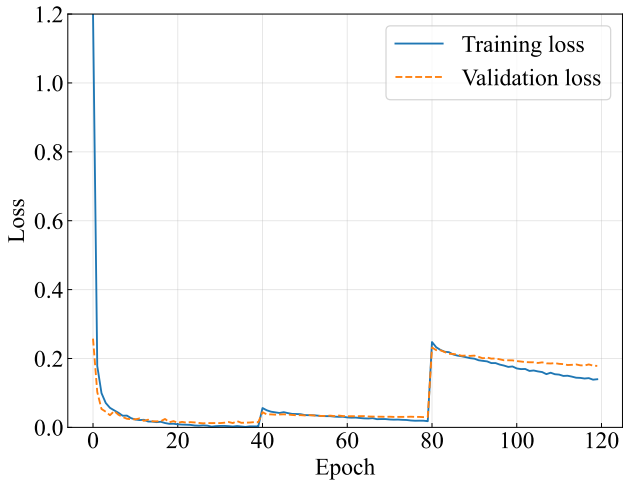


FIG. 5.3. Loss curves for the training and validation sets. During the first 40 epochs, we input samples with SNRs between [40, 50], followed by [30, 40] SNR samples in the subsequent 40 epochs, and [20, 30] SNR samples in the last 40 epochs.

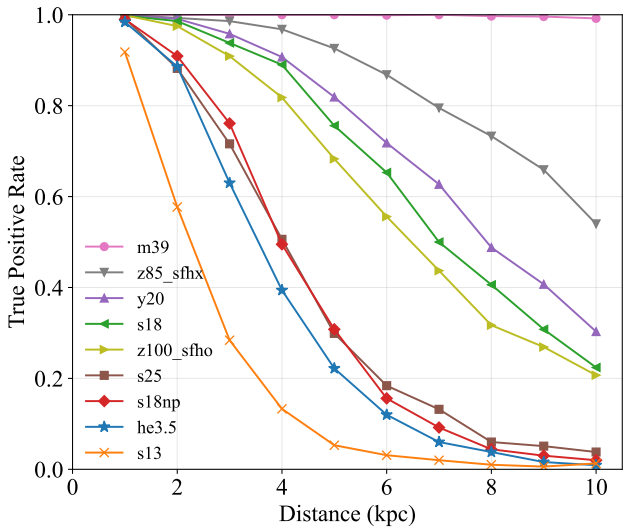
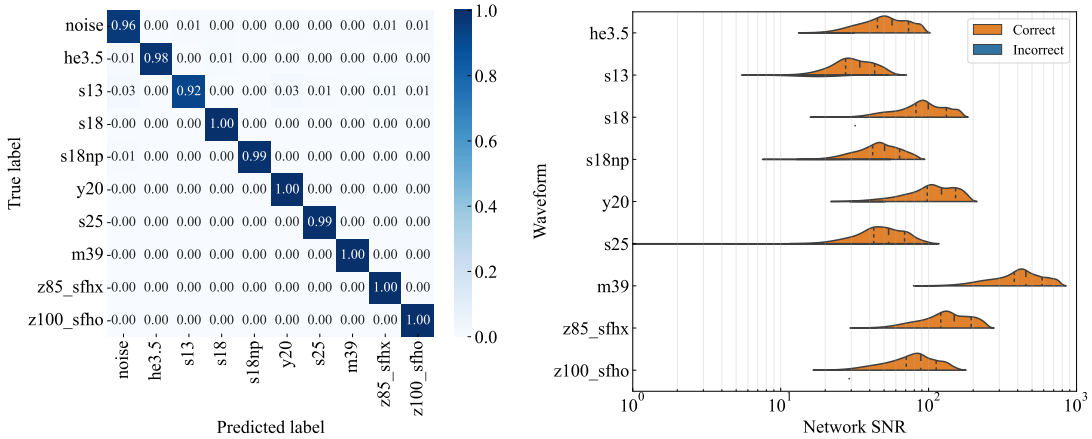


FIG. 5.4. True positive rate of each waveform in the test set against source distance.

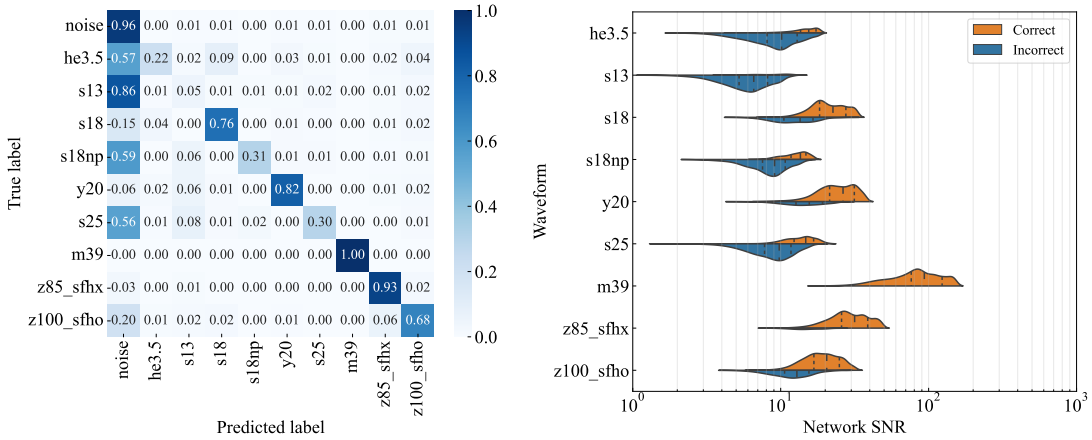
differences in the condition that we used O3 noise instead of O2 noise and performed ten-class classification instead of eight-class classification. In Fig. 5.4, we plot a true positive rate for each waveform in the test set against distance. A true positive rate, also known as the sensitivity of a class c, is defined as the ratio of the number of samples correctly classified into class c to the number of samples of class c in the test set. For signals from sources at 1 kpc, each waveform has a true positive rate greater than 90%, and this decreases monotonically with the distance of the source, having an average true positive rate of 26.1% at 10 kpc. For the `m39` waveform, because the amplitude of the strain is much larger than the others due to its rapid rotation and high explosion energy, the true positive rate for sources at 10 kpc is 99.2%.

The performance of a multiclass classifier is also expressed by a confusion matrix, which shows the number of samples classified into each class. Figure 5.5 plots the confusion matrices normalized for each class and the distribution of the network SNR for signals

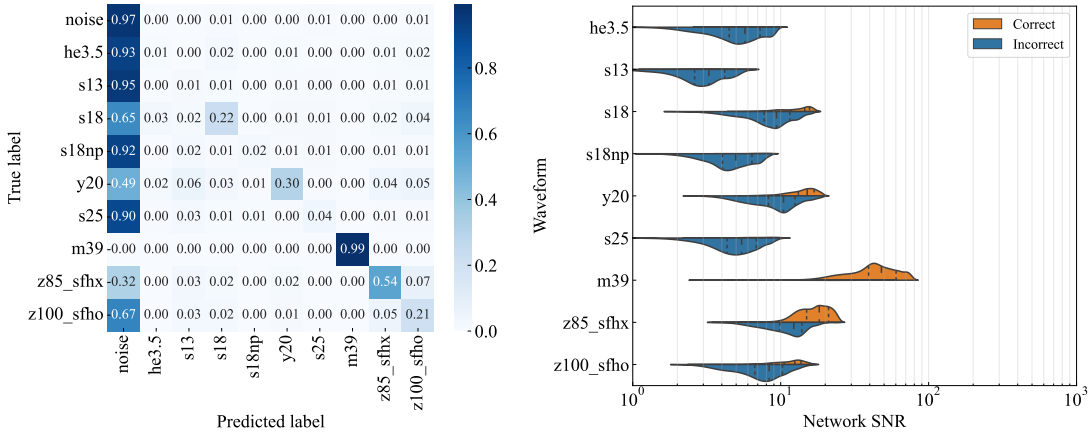
from sources at 1, 5, and 10 kpc. We can see from the confusion matrices that as the distance increases, the amplitude of the signal becomes smaller and the number of samples misclassified as noise increases. The accuracy for signals at 10 kpc is 33.2%, and our model cannot identify most of these signals, except for the `m39` waveforms, whose SNR is much higher than others with a median value of 47.9.



(a) 1 kpc



(b) 5 kpc



(c) 10 kpc

FIG. 5.5. Confusion matrices of the test set (left) and violin plots of network SNR of each waveform (right) from sources with distances of 1, 5, and 10 kpc.

5.3.2 Dimensionality reduction

Before implementing CAM techniques, we used the t-distributed stochastic neighbor embedding (t-SNE) [206] algorithm to see if the convolutional layers in the model can extract the features in the input to classify samples. The t-SNE algorithm is a dimensionality-reduction technique that minimizes the Kullback-Leibler divergence between two probability distributions: one representing pairwise similarities between data points in the original high-dimensional space and another representing pairwise similarities in a lower-dimensional space. In our CNN model, each sample is compressed into a vector with a length of 2112 before the dense layers. The t-SNE algorithm is used to map this vector into two-dimensional space to make it interpretable for humans. We visualize the dimensionally reduced feature maps of the test set, whose signals are coming from sources at 1 kpc, for which our model shows a good classification accuracy. The visualized data are shown in Fig. 5.6. We can clearly see that there are ten clusters in the data set and our model could extract meaningful features to classify these samples into ten classes. The fact that some signal samples are also found in the noise cluster and that `s13` samples are found in other clusters, especially in the noise cluster, is consistent with the results of the confusion matrix in Fig. 5.5(a).

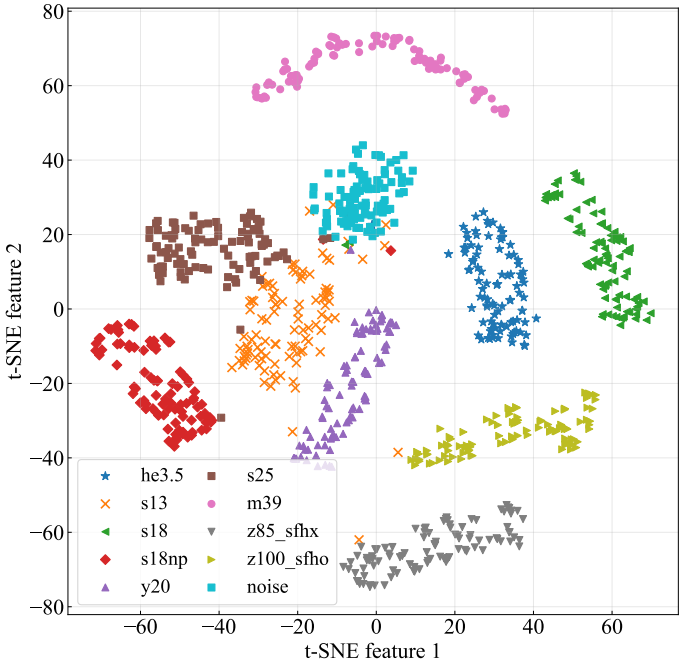


FIG. 5.6. Features of the test samples at 1 kpc extracted by CNN and mapped into two-dimensional space by the t-SNE algorithm.

5.3.3 Saliency maps

To quantitatively evaluate different CAM methods, we use two metrics—average drop and average increase [131]—which focus on the change in a model's score caused by the explanation map. An explanation map for a target class c is generated as element-wise

multiplication of a saliency map L^c with an original image X:

$$E^c = L^c \odot X. \tag{5.13}$$

Average drop measures the percentage decrease in a model's score for a target class c when inputting only the explanation map, instead of the original image. It is expressed as

$$\text{Average drop} = 100 \cdot \frac{1}{N} \sum_i \frac{\max(0, y_i^c - o_i^c)}{y_i^c}, \tag{5.14}$$

where y_i^c is the score for class c on the ith original image and o_i^c is the score on the explanation map. The lower this value, the more effective the visualization method, since the explanation map includes more of the relevant information for making a correct prediction.

Average increase measures the number of samples in the data set, and the model's confidence increases when providing only the explanation map as input. It is expressed as

$$\text{Average increase} = 100 \cdot \frac{1}{N} \sum_i \Theta(o_i^c - y_i^c), \tag{5.15}$$

where Θ is the Heaviside step function. Unlike the previous metric, the higher this value is, the more effective the visualization method will be because there are more samples that score higher when given the explanation map than when given the original image.

For the three visualization methods Grad-CAM, Grad-CAM++, and Score-CAM, the two metrics described above are computed using signals from sources at 1 kpc in the test set. The results are summarized in Table 5.1. Score-CAM shows the best results in both metrics, meaning that it is the best visualization technique for our model among the three CAM methods considered in this study. We also qualitatively compare these methods by visualizing some samples. One example is shown in Fig. 5.7. The input image is represented by a color image, with the red, green, and blue channels corresponding to the H1, L1, and V1 spectrograms, respectively. All three saliency maps take large values around the SASI mode around 100 Hz. At high frequencies, the Grad-CAM and Grad-CAM++ maps only take slightly larger values around 1 kHz, whereas Score-CAM has g-mode-like arch shapes around 1 kHz. This suggests that the visualization by Score-CAM captures more of the input features that are discriminative for the prediction.

TABLE. 5.1. Results for evaluation of the explanations generated by Grad-CAM, Grad-CAM++, and Score-CAM on the test set.

Method	Grad-CAM	Grad-CAM++	Score-CAM
Average Drop (%) (*Lower is better*)	30.70	17.58	**9.61**
Average Increase (%) (*Higher is better*)	1.30	1.40	**1.96**

As discussed above, we determine that the Score-CAM is the optimal method for generating saliency maps for our model. We produce saliency maps by Score-CAM for the inputs of each class, which can be seen in Fig. 5.8. In the input images, as in the previous figure, the red, green, and blue channels corresponding to the H1, L1, and V1

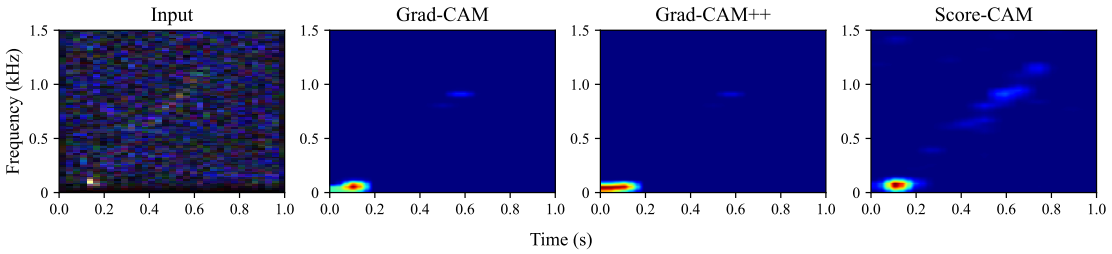


FIG. 5.7. Qualitative comparison of three CAM maps for the `s25` sample at 1 kpc.

data, respectively. This means, for example, that in the reddish image such as the `m39` sample in Fig. 5.8, the SNR at the H1 detector is smaller than that at the L1 and V1 detectors. All of the plotted signal samples are scaled to have an SNR of 40 and are correctly classified by our model. We plotted several CAM maps for noise samples, in addition to the one shown in this figure. However, the regions identified by the model for labeling the spectrograms as 'noise' do not exhibit a clear distribution; instead, they appear to be randomly distributed. In the `he3.5` and `s13` samples, we can see that the model focuses on the g-mode arch shape, especially in their low- and high-frequencies areas. In the `s18` and `y20` models, the models see all of the g-modes. In the `s18np` model, the CAM map indicates that the model considers not only the g-mode but also prompt convection and SASI. The `s25` model has SASI activity, but its amplitude is not too large, and the CAM map shows that the model's prediction is based on the prompt convection and the high-frequency g-mode. In the `m39`, `z85_sfhx`, and `z100_sfho` models, the CAM maps take large values at high frequencies in the g-mode. In addition, in the `z85_sfhx` and `z100_sfho` models, the low-frequency SASI mode, whose frequency increases with time, is also visible in the CAM maps. To summarize these outcomes, we found that the model looks at the g-mode in all signal waveforms, and also looks at SASI and prompt convection in some signal waveforms when classified.

Additionally, we plot saliency maps of the misclassified samples. Figure 5.9 shows a spectrogram of the `s25` signal sample and the Score-CAM map, which the model classified as `s18np`. An example spectrogram of the correct class `s18np` is also shown. The SNR of this signal is 85, which is quite large, and the g-mode and the prompt convection are visible, but there is a glitch in the strain at the L1 detector. The Score-CAM map shows that the model focuses on the prompt convection and the glitch, which are used to determine that the signal is `s18np`. Because of this glitch, the model predicted the signal as `s18np`, whose g-mode frequency increases in a shorter period of time. Another example is plotted in Fig. 5.10. This sample contains a `s13` signal with an SNR of 48, and there are no glitches, but the model classified it as `y20`. We can see a SASI-induced GW mode around 100 Hz from 0.2 to 0.4 s, but the Score-CAM map indicates that the model only looks at a portion of the g-mode and does not see the low-frequency mode.

From the misclassified samples and the Score-CAM maps, it is found that the performance of the model is sometimes affected by glitches, and does not fully take advantage of the characteristics of the signals. The former could be resolved by generating training sets that contain more glitches, and the latter could be resolved by using a time-frequency representation that is better able to reflect the various features of the CCSN signals.

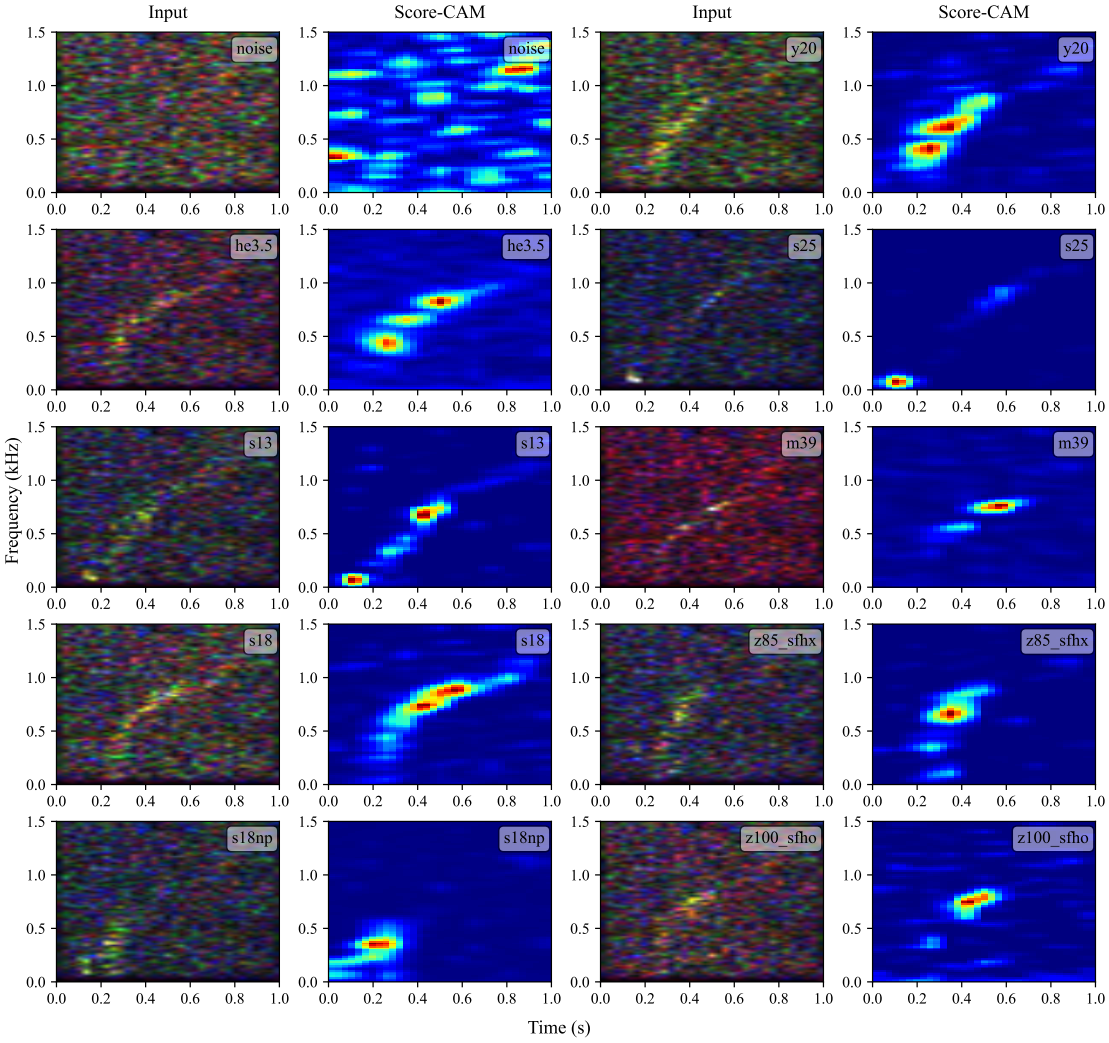


FIG. 5.8. Input spectrograms and Score-CAM maps of correctly classified samples. The SNR of each signal sample is 40.

5.4 Conclusions

In this study, we trained a two-dimensional CNN model to classify CCSN GW signals immersed in real noise of O3 observation data. Our model showed a comparable result to the previous study [152] for signals from sources with distances of 1 kpc. To interpret the trained model, we used the t-SNE algorithm and mapped the extracted features by the convolutional layers into a two-dimensional space. The dimension-reduced features show that the convolutional filters could extract meaningful features that are significant for classifying the signals. To gain insights into the decision-making process of the model, we applied the CAM technique to visualize the regions in the inputs that were influential to the predictions. Three methods—Grad-CAM, Grad-CAM++, and Score-CAM—were considered and we concluded that Score-CAM is the best for our model in terms of the average drop and average increase metrics. The Score-CAM maps of correctly classified signal samples revealed that the model's predictions were heavily affected by a part of the entire g-mode in the spectrogram of each signal. In some waveform models such as `s18np`, `s25`, `z85_sfho`, and `z100_sfho`, their CAM maps suggest that the prompt convection or

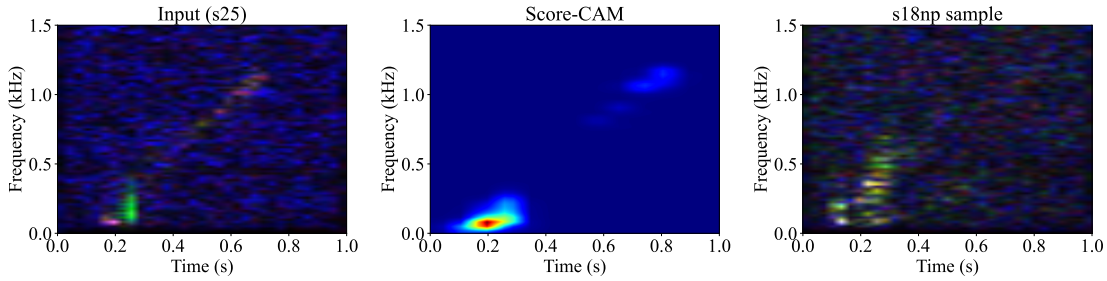


FIG. 5.9. `s25` sample classified as `s18np`. (a) Input spectrogram. The red, green, and blue channels correspond to the H1, L1, and V1 data, respectively. There is a glitch in the L1 data. (b) Score-CAM map. (c) Example spectrogram of a `s18np` sample.

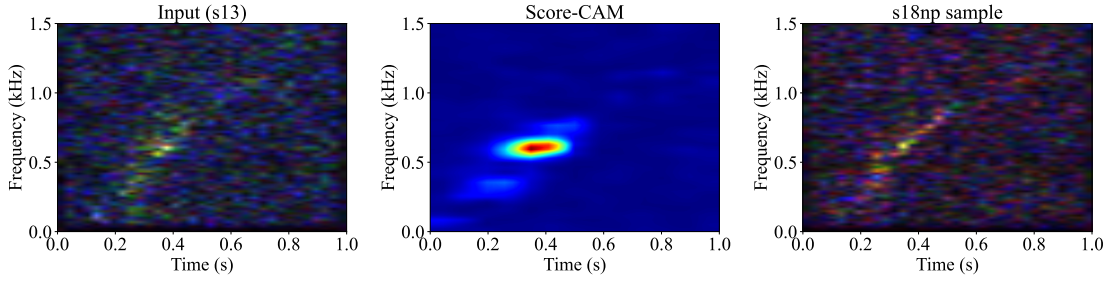


FIG. 5.10. `s13` sample classified as `y20`. (a) Input spectrogram. The red, green, and blue channels correspond to the H1, L1, and V1 data, respectively. (b) Score-CAM map. (c) Example spectrogram of a `y20` sample.

SASI-induced GW mode also affects the model's prediction.

It is important to note that $\sim 4\%$ of the pure noise test samples are identified as signal, which means that our model produces a false alarm every ~ 25 s, making it unsuitable as a detection pipeline. Since this study is the first to focus on the interpretability of CNN models in GW data analysis and serves as a first step in showcasing the effectiveness of the CAM techniques, we did not prioritize its viability as a detection methodology. To utilize machine learning models for a future detection pipeline, it is crucial to lower the false alarm rate. Using the CAM techniques can potentially enhance the efficacy of CNN models for this purpose.

In this analysis, a time-frequency map was created from the short-time Fourier transform, but its resolution is limited by the uncertainty relationship between time and frequency. In future studies, we would like to improve the accuracy of the CNN model by using methods such as the Hilbert-Huang transform [57], which can generate higher-resolution time-frequency maps, and to confirm that the CNN can also utilize several more GW modes to classify CCSN signals.

Chapter 6

Comparative Study of 1D and 2D CNN Models with Attribution Analysis for Gravitational Wave Detection from Compact Binary Coalescences

This chapter is a reprint of the paper [3] with minor modifications in formatting.

Abstract

Recent advancements in gravitational wave astronomy have seen the application of convolutional neural networks (CNNs) in signal detection from compact binary coalescences. This study presents a comparative analysis of two CNN architectures: one-dimensional (1D) and two-dimensional (2D) along with an ensemble model combining both. We trained these models to detect gravitational wave signals from binary black hole (BBH) mergers, neutron star-black hole (NSBH) mergers, and binary neutron star (BNS) mergers within real detector noise. Our investigation entailed a comprehensive evaluation of the detection performance of each model type across different signal classes. To understand the models' decision-making processes, we employed feature map visualization and attribution analysis. The findings revealed that while the 1D model showed superior performance in detecting BBH signals, the 2D model excelled in identifying NSBH and BNS signals. Notably, the ensemble model outperformed both individual models across all signal types, demonstrating enhanced detection capabilities. Additionally, input feature visualization indicated distinct areas of focus in the data for the 1D and 2D models, emphasizing the effectiveness of their combination.

6.1 Introduction

The era of gravitational wave (GW) astronomy was inaugurated with the first direct detection of GWs from a binary black hole (BBH) merger by the Advanced Laser Interferometer Gravitational-wave Observatory (Advanced LIGO) [5] in 2015 [6]. This

groundbreaking discovery was followed by the first joint observation of GWs and electromagnetic counterparts from a binary neutron star (BNS) merger, achieved by Advanced LIGO, Advanced Virgo [18], and other telescopes, paving the way for multi-messenger astronomy [21]. Over the course of three observing runs (O1, O2, and O3), 90 GW events from compact binary coalescences (CBCs) were reported [31–34]. These events included two neutron star-black hole (NSBH) mergers [30] and two BNS mergers [20, 207]. The detection of GWs, alongside electromagnetic waves and neutrinos from these mergers, is vital for understanding the physical properties of neutron star interiors, which are reflected in their equation of state. Now, with the commencement of the fourth observing run in May 2023, which includes the participation of KAGRA [19], expectations are high for more GW detections from binary systems with neutron stars.

Traditionally, GWs from CBC sources have been analyzed using the matched-filtering technique [37] with theoretical approximants, phenomenological models, and templates derived from numerical simulations [58, 105]. In this technique, the signal-to-noise ratio (SNR) is computed by correlating the detector's strain data with each template in a large bank that covers a wide parameter space, taking into account variations in source masses and/or spins. This method, however, can be computationally intensive, especially for complex GW signals that incorporate elements such as higher-order modes, precession, or orbital eccentricity. This complexity underscores the need for more efficient algorithms to manage the growing volume of GW data.

In response to this challenge, deep learning approaches, particularly convolutional neural networks (CNNs), have been increasingly applied in the GW field. These applications range from parameter estimation of CBC sources [48, 50] to sky localization [1, 163, 208] and classification of transient noises [51, 53, 209]. The effectiveness of CNNs in detecting GWs from BBH mergers was first demonstrated in 2018 by George and Huerta [44] and Gabbard *et al.* [134]. These initial studies have since been expanded to include more sophisticated models that use real detector noise and account for various signal complexities like the spin effect, precession, higher-order modes, or eccentricity [135, 136, 138, 140, 189, 210–219]. There are also some studies targeting BNS [45, 141–144, 220–222] or NSBH signals [146, 147, 223–225], which are more challenging than BBH signals due to their longer duration and smaller amplitude.

Two main types of CNNs have been employed in this research: one-dimensional (1D) CNNs, which process whitened time-series data, and two-dimensional (2D) CNNs, which analyze time-frequency maps. Although most studies have favored 1D CNNs [44, 45, 134, 135, 140, 141, 143, 147, 189, 210–212, 215–217, 220–222, 224, 225], a subset has opted for 2D CNNs [142, 144, 146, 213, 214, 219, 223]. 1D CNNs are preferred for their efficiency in not generating time-frequency maps, thereby reducing processing time. On the other hand, 2D CNNs excel at capturing the temporal evolution of GW frequencies in their input. For the analysis of GWs from core-collapse supernovae, Iess *et al.* [152] conducted a comparative study of 1D and 2D CNNs, alongside long short-term memory networks. Their approach involved combining these models by averaging their outputs. However, to our knowledge, a similar comprehensive comparison of various CNN architectures for CBC sources has not been extensively explored.

A common method for analyzing and interpreting CNNs involves pinpointing the segments of input data that significantly influence the model's predictions. This analysis can be performed using class activation mapping techniques [129] or by assessing the contribution of each input feature to the model's output. In our prior research [2], we applied CAM techniques to a CNN classifier designed for GWs from core-collapse supernovae. This investigation revealed that the model primarily focused on specific GW modes within the input spectrogram to make predictions.

In the current study, we train both 1D and 2D CNN models to detect and classify GWs from CBC sources. We then develop an ensemble model that combines these two CNN types. Our analysis includes a detailed comparison of the detection performance of these models across each type of CBC signals. To distinguish the different aspects that 1D and 2D models focus on within the input, we employ the integrated gradients technique [128]. This approach allows us to identify the influential regions in the input that guide the models' predictions, revealing distinct areas of focus between the 1D and 2D models.

This chapter is organized as follows. Section 6.2 details our data sets, the architecture of the CNN models, and the theoretical background of the CNN analysis methods. Section 6.3 presents the classification performance and a comprehensive analysis of our trained models. Finally, we conclude our findings in Sec. 6.4.

6.2 Method

Our CNN models are trained to classify strains at three detectors LIGO Hanford (H1), LIGO Livingston (L1), and Virgo (V1) into four distinct classes: BBH, NSBH, BNS, and pure noise. This section provides a detailed description of the data sets used for both training and testing our models. Following this, we describe the architecture and training procedures of the CNN models. Lastly, we address the dimensionality reduction technique implemented in our study, as well as the methodology employed for computing feature attribution, which are crucial for interpreting the models' decision-making processes.

6.2.1 Data set

To train and test our model, we used non-precessing CBC signals and injected them into noise obtained from O3 real data at H1, L1, and V1, which are available at the Gravitational Wave Open Science Center [203].

Signal and noise generation

To construct our data sets, non-precessing CBC signals were generated using the LIGO Algorithm Library Suite [226]. Specifically, BBH signals were simulated using the `SEOBNRv4` approximant [86], based on the effective-one-body method, while NSBH and BNS signals were generated using the `SpinTaylorT4` approximant [227], a time-domain post-Newtonian model incorporating spin effects. For BBH signals, component masses were uniformly sampled in the range of 5 to 80 $M_\odot$. NSBH signals had NS masses sampled between 1 and 2 $M_\odot$, and BH masses between 5 and 35 $M_\odot$. The component masses of BNS signals ranged uniformly from 1 to 2 $M_\odot$. The individual components have spins

aligned with the orbital angular momentum, uniformly distributed between 0 and 0.99. These waveforms were sampled at a rate of 4096 Hz. We used four-second data segments, with the merger event uniformly placed between 3.8 and 3.9 sec. Although NSBH and BNS signals are typically longer than 4 sec, we found this segment length is sufficient to discriminate between different classes. The use of shorter-segment signals also reduces the memory requirements for training models. The sky position of the source, defined by declination and right ascension, was randomly selected, and GW amplitude calculations were performed considering the antenna pattern functions and time delays across detectors. These computations utilized the PyCBC library [202].

For noise samples and background noise for signal samples, real strain data from GPS time 1238163456 to 1238659072 was used for the training set, 1238663168 to 1239162880 was used for the validation set, and 1239166976 to 1239875584 was used for the test set. Data around the GW event time reported in the GWTC-2.1 catalog [33] were excluded.

Preprocessing

After the signal samples were truncated to four-second segments, they were scaled based on the computed optimal matched-filter SNR, defined as

$$\rho = \sqrt{4 \int_{f_{\min}}^{f_{\max}} \mathrm{d}f \ \frac{|\tilde{h}(f)|^2}{S_n(f)}}, \tag{6.1}$$

where $\tilde{h}(f)$ is the Fourier transform of the truncated signal and $S_n(f)$ is the one-sided power spectral density of the noise, estimated using Welch's method [204]. The integration was performed from a cutoff frequency of 20 Hz up to the Nyquist frequency. The training and validation signals were scaled so that the network SNR of the three detectors, given by

$$\rho_{\mathrm{net}} = \sqrt{\rho_{\mathrm{H1}}^2 + \rho_{\mathrm{L1}}^2 + \rho_{\mathrm{V1}}^2}, \tag{6.2}$$

followed a uniform distribution between 8 and 24, while the SNRs of the test signals ranged from 3 to 24. After each signal was injected in noise, we whitened the sample in frequency domain using the power spectral density. For the input to the 2D CNN model, we generated time-frequency maps using the Q transform [113] of the whitened samples, defined by

$$X(\tau, \phi, Q) = \int_{-\infty}^{\infty} \mathrm{d}f \ \tilde{x}(f + \phi)\tilde{w}^*(f, \phi, Q)e^{-2\pi i f \tau}, \tag{6.3}$$

where Q is the quality factor, and the Connes window functions is used as the window function [114].

The final data set comprised 408,000 training samples, 408,000 validation samples, and 528,000 test samples. Each data set had an equal distribution of 25% BBH, 25% NSBH, 25% BNS, and 25% pure noise samples. Representative samples from each class in the training set are displayed in Fig. 6.1.

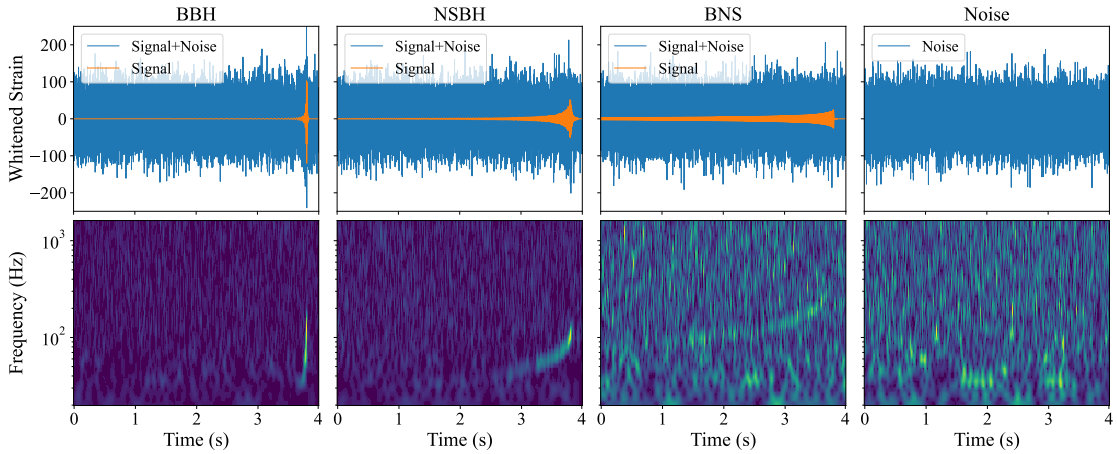


FIG. 6.1. Example strain data at H1 detector in the training set. The upper figures show the whitened time-series data used as input to the 1D model, and the lower figures show the time-frequency maps for the 2D model. The component masses of the BBH signal are 51.3 $M_\odot$ and 50.9 $M_\odot$, whereas in the NSBH sample, the respective masses are 33.3 $M_\odot$ and 1.83 $M_\odot$, and the masses of the BNS sample are 1.54 $M_\odot$ and 1.40 $M_\odot$. The single-detector SNR of each signal is 15, and the merger time is fixed at 3.9 s.

6.2.2 Model

1D CNN

One-dimensional CNN consists of 1D convolutional filters. Let x_i^c be the ith value of the cth channel of the input series and y_i^m be the ith value of the mth channel of the output. The output of a 1D convolutional filter is given by

$$y_i^m = \sum_{c=0}^{C-1} \sum_{k=0}^{K-1} w_k^m x_{i+k}^c + b^m, \tag{6.4}$$

where K is the kernel size and C is the number of input channels. Weight parameters w and bias parameters b are learned during training processes.

Our 1D CNN model takes a three-channel whitened time series at H1, L1, and V1 as input. Our implementation uses a 54-layer deep residual neural network (ResNet-54) which was proposed in Ref. [140]. ResNet is a type of deep CNN architecture that uses residual blocks to address the vanishing gradient problem commonly encountered in deep networks. It achieves this by adding *skip connections* between layers, enabling the network to learn residual functions and make training deep networks more efficient [127]. Details of the ResNet-54 model architecture can be found in Ref. [140].

For input normalization, a deep adaptive input normalization layer [228] is employed in this analysis as used in Ref. [140] to address non-stationary noise that appears in real detector noise. In this layer, unlike conventional normalization, shifting and scaling parameters for normalizing the input are optimized during the training. Including the parameters in the normalization layer, the model has a total of 1,935,698 trainable parameters.

The employed architecture is one of the state-of-the-art models in BBH detection that surpassed the matched-filtering pipeline in a specific condition [140]. Alternatively, we also employed the CNN architecture used in Ref. [147] designed to detect all types of CBC

signals; however, it did not show a better performance than the ResNet-54 model in our data sets.

2D CNN

Two-dimensional CNN consists of 2D convolutional filters. Let $x^c_{i,j}$ be the (i,j) component of the cth channel of the input image and $y^m_{i,j}$ be the (i,j) component of the cth channel of the output. The output of a 2D convolutional filter is given by

$$y^m_{i,j} = \sum_{c=0}^{C-1} \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} w^m_{k,l} x^c_{i+k,j+l} + b^m \tag{6.5}$$

where (K, L) is the kernel sizes and C is the number of input channels. Weight parameters w and bias parameters b are learned during training processes.

The ResNet-50 model [127], a variant of ResNet with proven efficacy in image recognition tasks, forms the basis of our 2D CNN. This model processes three-channel images of time-frequency maps and includes 23,508,548 trainable parameters. We adopt this model because it is one of the most widely used 2D CNNs in GW signal detection and has a similar number of layers to our 1D model. Its efficiency has been validated in previous studies [142, 146, 223].

Ensemble model

In our ensemble approach, we combine the outputs of the 1D and 2D CNN models to enhance predictive performance. This is achieved by first training a fully connected neural network, which takes as input a concatenated vector of features extracted from the trained 1D and 2D models. The input vector, with a dimension of 10240, is processed through a hidden layer of 200 units, outputting a four-dimensional vector. The network incorporates a Leaky ReLU layer [229] and a dropout layer [230] with a 0.25 dropout rate for regularization. The ensemble network comprises 2,049,004 trainable parameters.

For the final model output, we employ a weighted average of the predictions from the 1D, 2D, and ensemble network. The weights, optimized for accuracy on the validation set, are set at 0.4 for each of the 1D and 2D CNNs and 0.2 for the ensemble network. The ensemble model is illustrated in Fig. 6.2.

Training process

Both the 1D and 2D CNN models, as well as the ensemble network, were developed using the PyTorch library [231] and trained on four NVIDIA Tesla V100 GPUs. All models were trained using categorical cross entropy as the loss function and Adam optimizer [123] with an initial learning rate of 10^{-3}. The learning rate was controlled by PyTorch's `ReduceLROnPlateau` method. During the training of the 1D and 2D CNNs, we implemented the curriculum learning technique [205]. This method involves initially training with high SNR samples and progressively incorporating lower SNR samples, thereby improving learning efficiency and model performance.

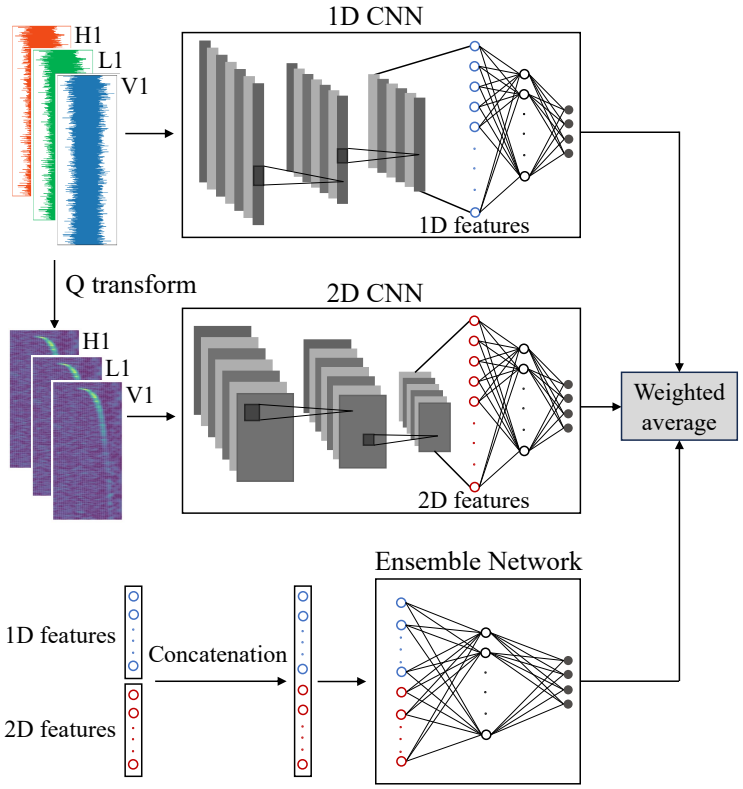


FIG. 6.2. Illustration of the ensemble model. The final output is the weighted average of the outputs from the 1D CNN, 2D CNN, and ensemble network.

The 1D CNN underwent 300 epochs of training with a mini-batch size of 1024. In contrast, the 2D CNN was trained for 45 epochs using a mini-batch size of 36. The ensemble network's training lasted for 25 epochs with a mini-batch size of 256.

6.2.3 t-distributed stochastic neighbor embedding

To understand the ability of 1D and 2D convolutional filters to extract meaningful features from the input for classification, we analyzed feature maps from the final convolutional layer. Given the high-dimensional nature of these feature maps, we utilized the t-distributed stochastic neighbor embedding (t-SNE) technique [206] for dimensionality reduction.

In the original space, the conditional probability $p_{j|i}$ that a data point x_i would pick a data point x_j as its neighbor is modeled as a Gaussian distribution centered at x_i, defined as

$$p_{j|i} = \frac{\exp\left(-\|x_i - x_j\|^2/2\sigma_i^2\right)}{\sum_{k\neq i}\exp\left(-\|x_i - x_k\|^2/2\sigma_i^2\right)}, \tag{6.6}$$

where σ_i is the standard deviation. We define the joint probability p_{ij} as the symmetrized conditional probabilities, which can be expressed as $p_{ij} = (p_{j|i} + p_{i|j})/2n$, where n is the number of data points. In the low-dimensional space, the Student t distribution with one

degree of freedom, defined by

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq l}(1 + \|y_k - y_l\|^2)^{-1}}, \tag{6.7}$$

is used to quantify the similarity of data points. The optimal low-dimensional representations are obtained by minimizing the Kullback-Leibler divergence of the distributions p_{ij} and q_{ij}, given by

$$C = \sum_{i,j} p_{ij} \log \frac{p_{ij}}{q_{ij}}. \tag{6.8}$$

The value of σ_i in Eq. (6.6) is determined by selecting the hyperparameter called perplexity, which is defined as 2 to the power of the Shannon entropy. The perplexity can be interpreted as a measure of the number of valid neighbors, and typical values are between 5 and 50 [206]. We set the perplexity at 25.

6.2.4 Integrated gradients

To discern which aspects of the inputs significantly influence the predictions in our trained 1D and 2D models, we employed the integrated gradients method [128]. While class activation mapping techniques [129, 130] are commonly used for such analysis, they often yield low-resolution saliency maps, especially in deep models. To circumvent this limitation, the integrated gradients method provides high-resolution feature attribution maps, proving advantageous for our analysis.

The integrated gradients method is grounded in two axioms that attribution methods should satisfy: (i) *sensitivity*, where any difference in one feature between the input and the baseline resulting in different predictions should receive a non-zero attribution, and (ii) *implementation invariance*, where the attributions for two functionally equivalent networks should be always identical.

Consider a function F that represents a network and let x be the input and x' be the baseline input. The feature attribution map is calculated by examining the path from the baseline x' to the input x and accumulating the network's gradients along this path. A point on this path can be expressed as $x' + \alpha(x - x')$ where α varies from 0 to 1. The integrated gradients along the ith dimension for an input x are defined as

$$\mathrm{IG}_i(x) = (x_i - x'_i) \int_0^1 \mathrm{d}\alpha \; \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i}. \tag{6.9}$$

In practice, this integration is approximated using the Riemann sum, described as

$$\mathrm{IG}_i(x) \approx (x_i - x'_i) \sum_{k=1}^{N} \frac{\partial F(x' + \frac{k}{N}(x - x'))}{\partial x_i} \frac{1}{N}. \tag{6.10}$$

Here N represents the number of interpolation steps. For accurately approximating the integral, a step size ranging from 20 to 300 is typically effective [128]. In our implementation, we chose $N = 30$ steps. The Captum library [232] was utilized to compute the attribution maps using the integrated gradients method.

6.3 Results and discussion

6.3.1 Model performance

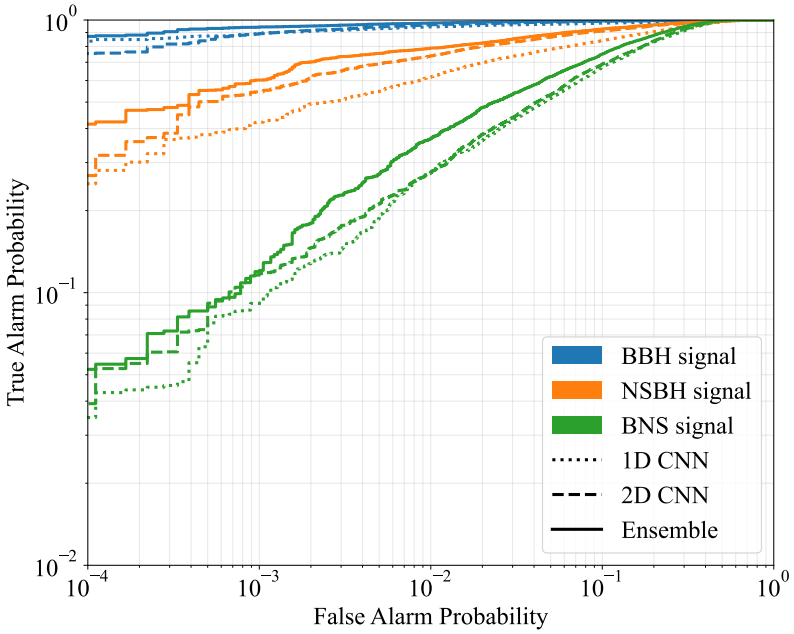


FIG. 6.3. ROC curves of the three models for BBH, NSBH, and BNS signals at a fixed network SNR of 8.

To evaluate the performance of our three models (1D, 2D, and the ensemble model), we first examined the receiver operating characteristic (ROC) curves for each signal type. The ROC curve plots the true alarm probability against the false alarm probability at various classification thresholds. As depicted in Fig. 6.3, the ROC curves for each type of GW signal at a fixed network SNR of 8 show distinctive sensitivities.

It was observed that all models exhibited the highest sensitivity to BBH signals, followed by NSBH and BNS signals. This trend aligns with expectations considering the relative amplitude of each signal type. Notably, the ensemble model demonstrated superior performance across all signal types. For BBH signals, the performance ranking was ensemble, followed by the 1D and then the 2D model. In contrast, for NSBH and BNS signals, the 2D model outperformed the 1D model. This variation in performance can be attributed to the transient nature of BBH signals, which are more effectively captured by the 1D convolution in time-series data. Conversely, the smaller amplitudes of NSBH and BNS signals, which are more challenging to identify in time-series data, render the 2D model more effective. This difference highlights the effectiveness of combining the 1D and 2D models.

We further calculated the detection sensitivity for each signal type as a function of network SNR. Figure 6.4 shows the sensitivity curves for the three models at a fixed false alarm probability of 0.001. The 1D model's sensitivity is on par with that reported in Ref. [147], where the model was trained using single-detector input. Their model's sensitivity saturates at a single-detector SNR of $\rho_{\mathrm{L1}} \geq 8$ for BBH signals, at $\rho_{\mathrm{L1}} \geq 10$ for NSBH signals, and at $\rho_{\mathrm{L1}} \geq 13$ for BNS signals. However, our 1D model reaches

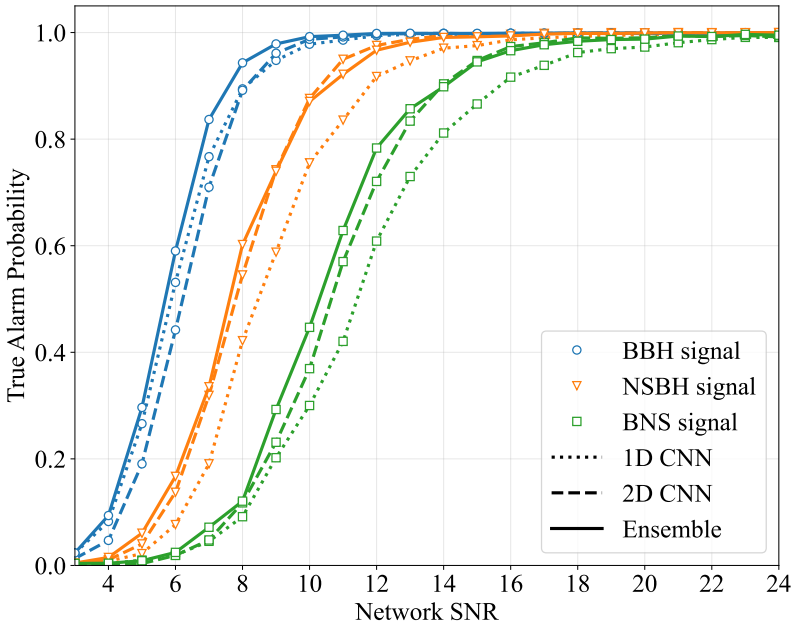


FIG. 6.4. Sensitivity curves of the three models for BBH, NSBH, and BNS signals at a fixed false alarm probability of 0.001.

saturation for BBH signals at $\rho_{\text{net}} \geq 12$, for NSBH signals at $\rho_{\text{net}} \geq 17$, and for BNS signals at $\rho_{\text{net}} \geq 22$. Given that the network SNR of three detectors is roughly $\sqrt{3}$ times that of a single-detector SNR, the performance of our 1D model is consistent with their model. The ensemble model further enhances this performance, lowering the saturation SNRs to 10 for BBH signals, 14 for NSBH signals, and 21 for BNS signals.

6.3.2 Feature map

We extract feature maps from the final convolutional layers. Since these vectors are fed into fully connected layers to make predictions, these feature maps represent the characteristics of each class. Figures 6.5 and 6.6 display the t-SNE projections of the feature maps for the trained 1D and 2D models, respectively. For these visualizations, we randomly selected 200 samples from the test set. In the figures, the size of each marker representing a signal sample is proportional to its SNR. Smaller markers indicate lower SNR signals, while larger markers correspond to higher SNR signals.

Figure 6.5 shows that the high-SNR BBH and NSBH samples are distinctly separated from the noise cluster, indicating effective classification of these signals by the 1D model. However, BNS samples are observed to be closer to the noise cluster, suggesting less clear differentiation for this signal type. Low-SNR signals across all types are more diffusely distributed within the noise cluster. In contrast, Fig. 6.6 indicates that the 2D model has an improved ability to separate not only the BBH and NSBH signals but also the BNS signals from the noise cluster. This indicates that the 2D model may be more proficient at identifying features of BNS signals compared to the 1D model.

Additionally, we explored embedding the feature maps into a three-dimensional space. However, this analysis revealed similar features to those observed in the two-dimensional embeddings.

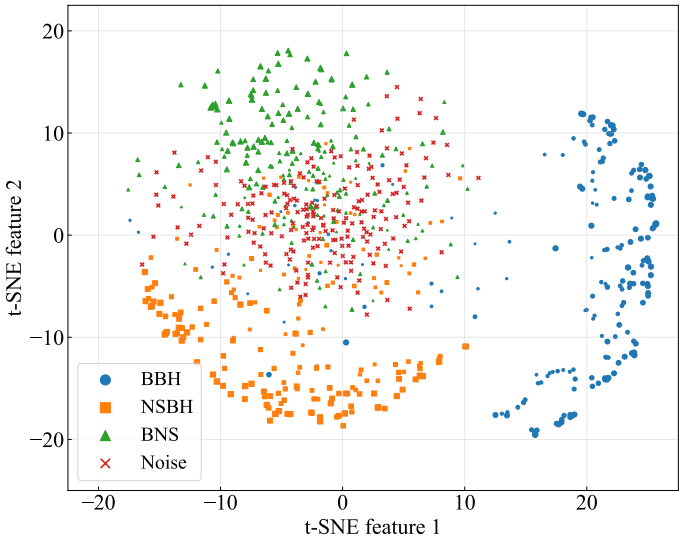


FIG. 6.5. Two-dimensional representations of feature maps of the 1D model by t-SNE. The marker size of signal sample corresponds to the SNR.

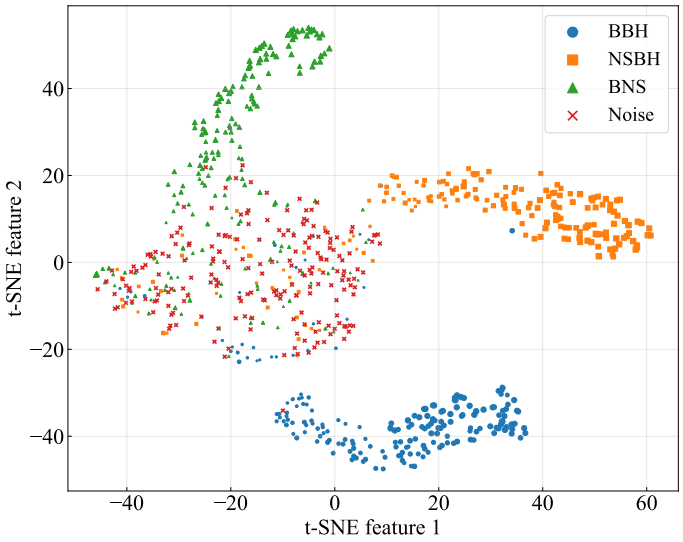


FIG. 6.6. Two-dimensional representations of feature maps of the 2D model by t-SNE. The marker size of signal sample corresponds to the SNR.

6.3.3 Attribution map

To understand how input features contribute to model predictions, we generated attribution maps for each type of signal using the integrated gradients method. For this analysis, a single signal was randomly selected from the test set, and ten distinct noise samples were added to create ten different input samples. Attribution maps were produced for each input sample, and their average was computed to discern universal characteristics of the attribution.

Figures 6.7, 6.8, and 6.9 present the attribution maps for a BBH, NSBH, and BNS signal, respectively, as identified by the 1D and 2D models. Each signal had a fixed

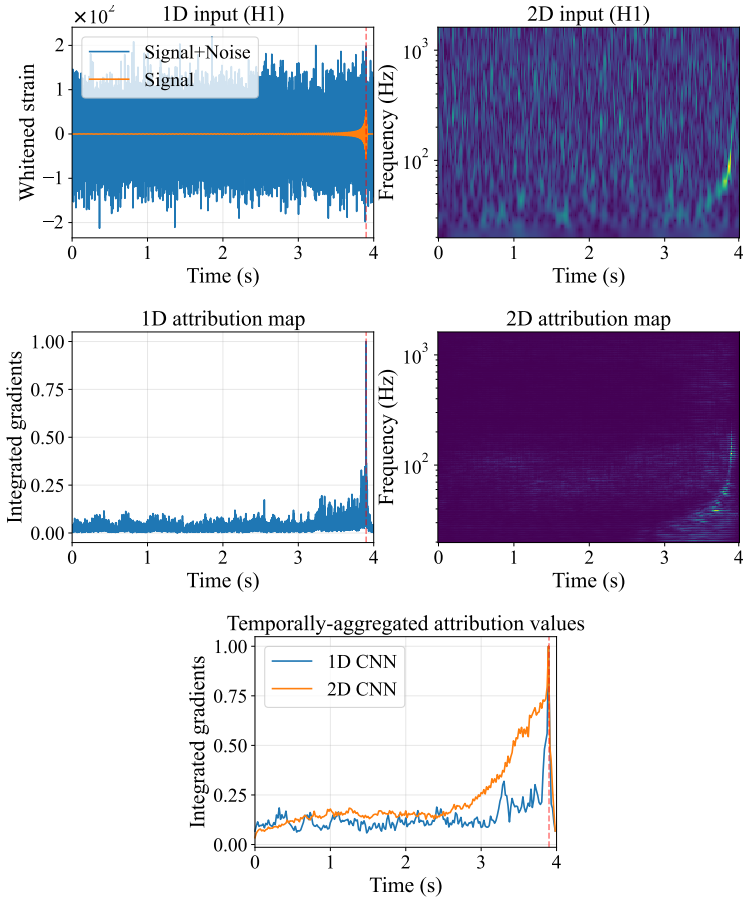


FIG. 6.7. Attribution maps of the 1D and 2D models for a BBH signal and corresponding input samples at H1 detector. The red dashed line shows the time of coalescence.

network SNR of 20, and the attribution maps for both models were normalized to a [0, 1] range. The bottom plots in Figs. 6.7, 6.8, and 6.9 show the values of integrated gradients summed over all frequencies for each time bin for 1D and 2D model, respectively. The values are normalized to a maximum integrated gradients of one for each model.

In Fig. 6.7, the 1D model shows significant contributions from data at the coalescence time of the BBH signal. In contrast, the 2D model's attribution map indicates that the 2D model focuses on the entire inspiral signal. Both the 1D and 2D models exhibit similar characteristics when integrated-gradients values are temporally aggregated. However, the 2D model sees data at more broader time frame than the 1D model.

As for the NSBH signal, the 1D model exhibits multiple peaks in the integrated gradients values before the coalescence, with the peak values progressively increasing, shown in Fig. 6.8. Since the model not only detects the signal but also classifies it into three classes, the data prior to the time of coalescence seem to be more significant than the data at the time of coalescence for determining that the signal is NSBH, not BBH. In the 2D model, the feature contribution of the entire inspiral is large, as in the case of the BBH signal. Similar characteristics of the NSBH sample are seen in the BNS sample in Fig. 6.9, but in the case of 1D attribution map of the BNS sample, peaks are also seen at earlier times and the overall values of the integrated gradients are generally identical. This indicates that the model focuses on various parts of the input time series data, which is reasonable

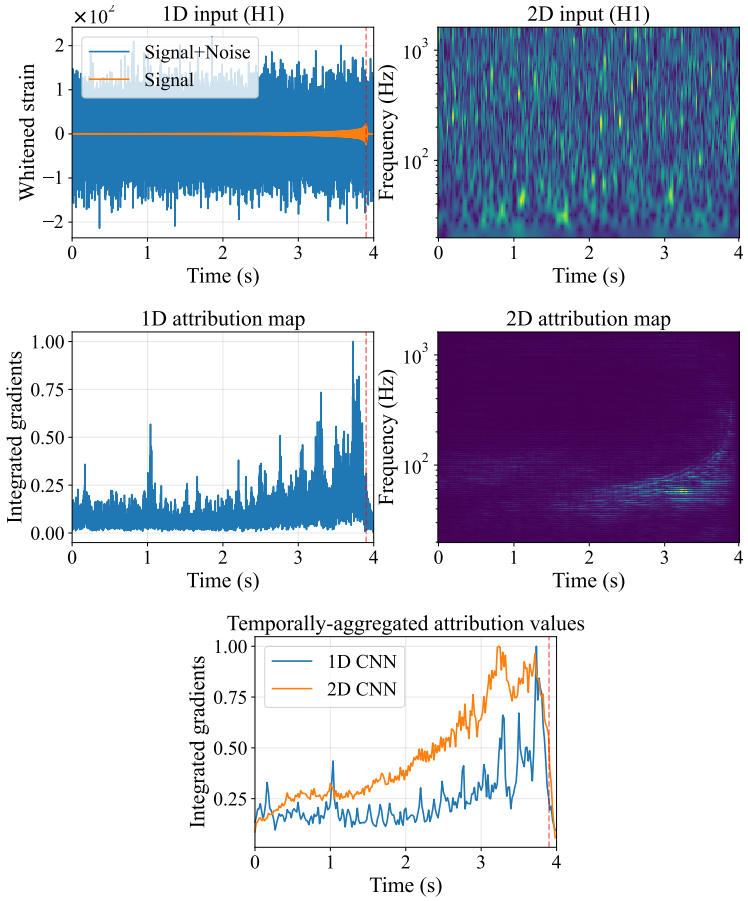


FIG. 6.8. Attribution maps of the 1D and 2D models for a NSBH signal and corresponding input samples at H1 detector. The red dashed line shows the time of coalescence.

because the inspiral signal is longer than the other signals. The 2D attribution map of the BNS sample shows that the 2D model accurately captures the BNS chirp signal on the input spectrogram, and it demonstrates the consistent performance of the 2D model for BNS events. The temporally aggregated attribution maps have similar characteristics for each signal type for the 1D and 2D models, but the 2D model shows a greater emphasis on longer signal durations than the 1D model.

In summary, from the attribution maps, we observe that the 1D model places greater emphasis on the time preceding coalescence, especially as the waveform lengthens, with significant contributions from specific moments in the inspiral phase. Conversely, the 2D model assesses the entire chirp waveform in the spectrogram, classifying based on the shape of the chirp, i.e., the temporal evolution of its frequency.

6.4 Conclusions

In this study, we explored the application of three distinct models: a 1D CNN, a 2D CNN, and an ensemble model for detecting and classifying GWs from CBC sources. The 1D model, trained on whitened time-series data, excelled in identifying BBH signals, while the 2D model, trained on Q-transformed spectrograms, showed superior performance

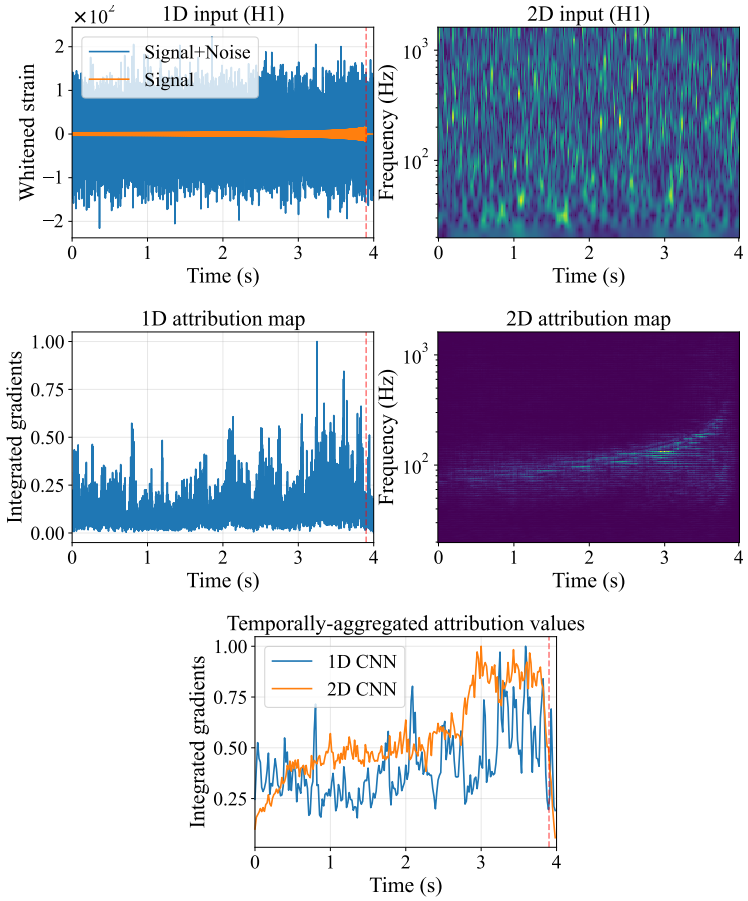


FIG. 6.9. Attribution maps of the 1D and 2D models for a BNS signal and corresponding input samples at H1 detector. The red dashed line shows the time of coalescence.

with NSBH and BNS signals. Overall, the ensemble model demonstrated the most robust classification capability across all signal types.

The effectiveness of combining 1D and 2D models was further reinforced through feature map visualization using the t-SNE technique and attribution map analysis via the integrated gradients method. We observed that the 1D model tends to focus on data preceding the merger time, especially as signal duration increases. In contrast, the 2D model scrutinizes the entire chirp waveform, capturing the intricacies of GW signals more comprehensively. These differences in focus and performance between the models highlight the benefits of their integration.

While our study presents results based on a specific selection of architectures for 1D and 2D CNNs, it is important to recognize that the field offers a wide variety of CNN architectures. Future research exploring additional architectures may provide a more comprehensive understanding and validation of the conclusions drawn in this study.

We discussed the performance of the models by fixing a false alarm probability at 0.001, but given that each data sample is 4 s long, this would result in roughly one false positive every hour, making our method insufficient for real application. Efficiency could be enhanced by incorporating a subsequent model, such as a binary classifier differentiating BBH signals from noise, to further reduce the false alarm probability. Validation is also required to address unbalanced data, considering the source population.

Our classification models hold potential for analyzing long continuous data through a sliding window approach. Although our models were trained to identify GW signals occurring between 3.8 and 3.9 s within four-second segments, sliding input window with a step size, for example, of 0.1 s, allows us to detect signals at any time point in principle. However, this approach may lead to encountering multiple triggers within a single event, requiring further tuning, which will be addressed in future studies.

Chapter 7

Using Hilbert-Huang Transform for Parameter Estimation of Proto-Neutron Stars from Their Gravitational Waves

Abstract

Core-collapse supernovae are potential multi-messenger events detectable by current and future gravitational-wave detectors. The gravitational-wave signals emitted during these events are expected to provide insights into the explosion mechanism and the internal structures of neutron stars. In recent years, several studies have empirically derived the relationship between gravitational-wave modes originating from the oscillations of proto-neutron stars and the parameters of these stars. This study applies the Hilbert-Huang transform to extract the frequencies of these modes and infer the physical properties of proto-neutron stars. The results exhibit comparable accuracy to a short-time Fourier transform-based estimation, highlighting the potential of this approach as a new method to extract physical information from gravitational-wave signals from core-collapse supernovae.

7.1 Introduction

Nearby core-collapse supernova (CCSN) explosions are expected sources for ground-based gravitational-wave (GW) detectors. While the explosion mechanism is not yet fully elucidated, most stars are believed to undergo neutrino-driven explosions [91, 233]. Current detectors, such as Advanced LIGO, are predicted to detect explosions occurring within our galaxy, up to a distance of approximately 10 kpc [188]. The Einstein Telescope [66] is one of the proposed ground-based detectors, consisting of three interferometers with 10 km arms arranged in a triangular formation. Such future detectors are anticipated to enable the detection of supernova events from galaxies such as the Large Magellanic Cloud.

Over the past decade, significant progress has been made in the multi-dimensional numerical simulation of CCSN explosions. In the neutrino-driven signals, the primary source of GW emission is the g-mode, which represents the oscillation of proto-neutron

stars, and the non-spherical flow of hot matter. The frequencies of the former mode progressively increase from around 100 Hz up to several kHz, while the latter represents a low-frequency mode around 100 Hz. The temporal evolution of these modes carries valuable physical information about proto-neutron stars, and their observation contributes to the field of asteroseismology.

In recent years, a relationship between the frequency of g-modes and the mass and radius of proto-neutron stars has been investigated through numerical simulations. In the study by Torres-Forné *et al.* [234, 235], relationships between each $g/f/p$-mode and the mass and radius of the proto-neutron stars, as well as the shock mass and radius, were calculated. These relationships are universal, as they do not depend on the equation of state or the progenitor mass. Similar universal relationships have also been explored in Ref. [236]. Using these relationships, the parameter estimation of proto-neutron stars has been performed in Refs. [237, 238] by using short-time Fourier transform (STFT) to extract g-mode frequencies.

To infer the physical properties of GW sources, accurate extraction of GW frequencies is essential. For this purpose, the use of Hilbert-Huang transform [56] has been studied in the GW field. Compared to the conventional STFT, time-frequency maps generated by the Hilbert-Huang transform are not affected by the trade-off relationship between temporal duration and bandwidth, resulting in high-resolutions both in time and frequency. Several studies showcased the effectiveness of the Hilbert-Huang transform in determining the frequencies of GWs, including those from post-merger binary neutron star coalescences [173, 182], and the mode induced by the standing accretion shock instability (SASI) in a CCSN [57].

In this study, we conduct a basic investigation using the Hilbert-Huang transform to enhance the estimation of M/R^2 from the g-mode GW signal, where M and R denote the mass and radius of a proto-neutron star. We use a simulated GW signal from a CCSN explosion and inject it into simulated Gaussian noise of the Einstein Telescope. Initially, we apply the method to the raw GW signal, and subsequently, we examine noise-added signals from various distances, conducting a comparative evaluation of accuracy with an approach based on the STFT.

The remainder of this chapter is structured as follows. In Sec. 7.2, we describe the simulation data used in this analysis. In Sec. 7.3, we explain our Hilbert-Huang transform-based parameter estimation method. Our results are presented in Sec. 7.4. We discuss the results in Sec. 7.5 and conclude this chapter in Sec. 7.6.

7.2 Simulation data

We use the GW signal of the `he3.5` model derived from a three-dimensional numerical simulation by Powell and Müller [194]. The progenitor is an ultra-stripped star evolved from a helium core with an initial mass of 3.5 $M_\odot$. An ultra-stripped star is a stellar remnant resulting from the extreme stripping of outer layers in a binary star system [239]. The simulation was performed with the neutrino hydrodynamics code CoCoNuT-FMT [198]. The shock is revived at 0.4 s and the simulation was stopped at 0.7 s after core bounce. The GW signal is dominated by the ${}^2g_2$-mode with a peak frequency of $\sim$ 800 Hz. The

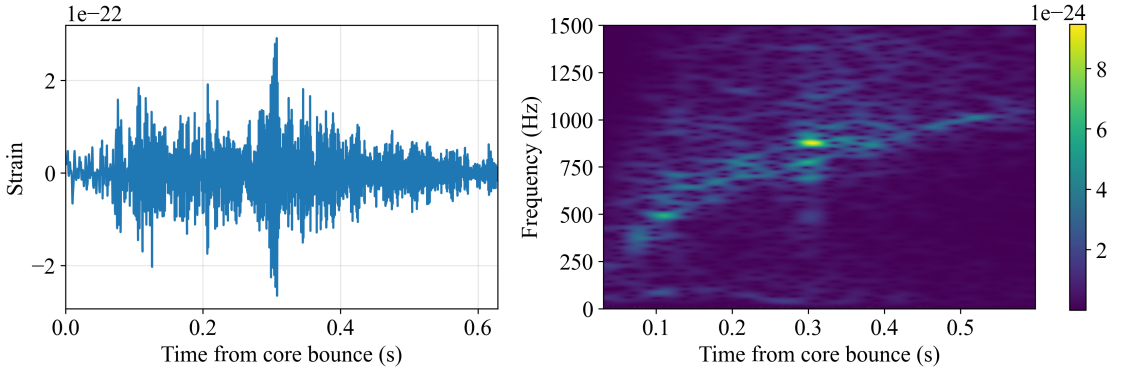


FIG. 7.1. The plus mode GW signal of the `he3.5` model in time-domain (left) and in time-frequency domain (right). The time-frequency map is produced using short-time Fourier transform. The source is placed at 8 kpc, and the signal is observed from the equatorial plane.

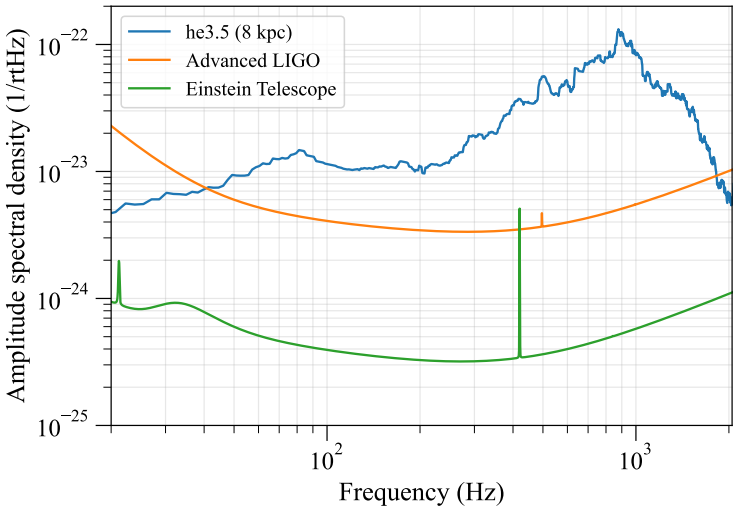


FIG. 7.2. Amplitude spectral density of the `he3.5` GW signal from 8 kpc and the design sensitivities of Advanced LIGO and the Einstein Telescope. The design sensivitiy of Advanced LIGO is obtained from Ref. [240]. We use the design sensitivity of the 'ET-D' configuration [241] for the Einstein Telescope.

model does not exhibit low-frequency GW emission from the SASI, due to the small core mass and the rapid drop in the accretion rate [194]. The plus mode of the GW signal is plotted in Fig. 7.1. In Fig. 7.2, we plot the amplitude spectral density of the characteristic strain at 8 kpc, along with the design sensitivity curves for Advanced LIGO and the Einstein Telescope.

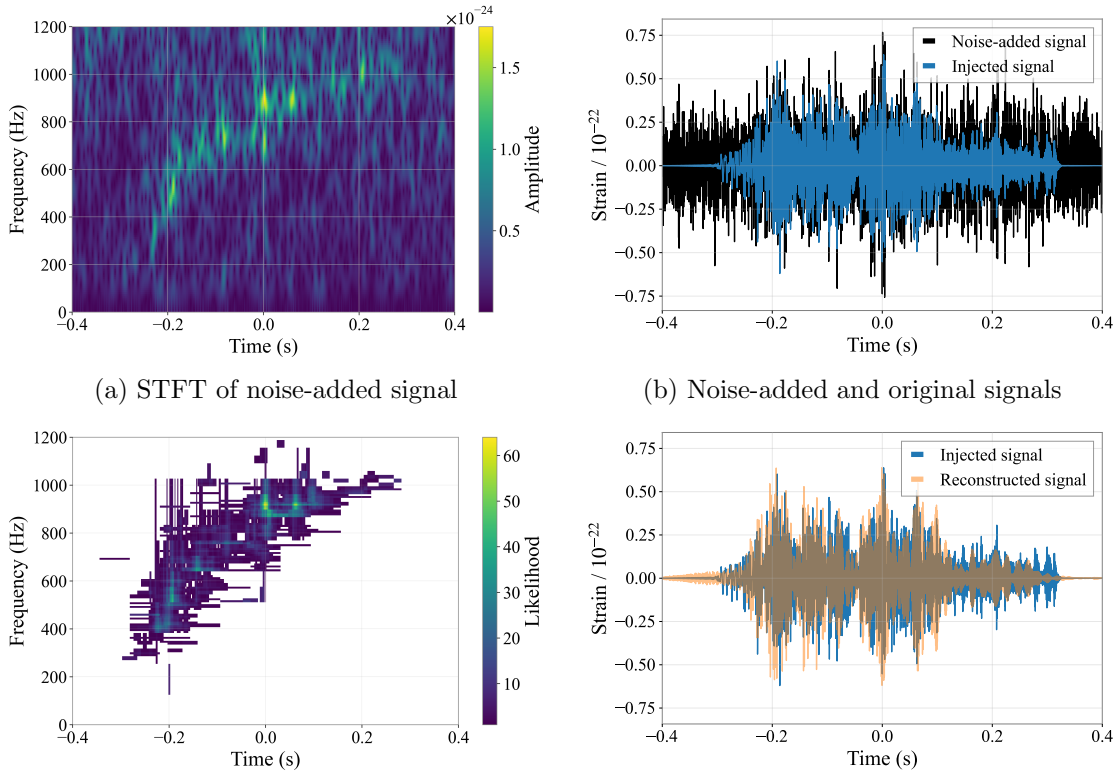
For the remainder of the analysis, the observer direction is fixed in the equatorial plane, and the sky position is set at the Galactic Center. We sample GW signals and detector noises from the Einstein Telescope at a sampling rate of 4096 Hz and apply a 100-1200 Hz band-pass filter. The selected frequency band is reasonable considering the g-mode frequency shown in the spectrogram in Fig. 7.1. Additionally, we apply a notch filter around 420 Hz to eliminate noise, which is represented as a sensitivity peak in Fig. 7.2. We set the time origin when the GW amplitude is at its maximum.

7.3 Parameter estimation method

7.3.1 Signal reconstruction

Since the Hilbert-Huang transform is susceptible to degradation from noise, we perform signal reconstruction before applying the Hilbert-Huang transform to the GW signal injected into detector noise. For signal reconstruction, we employ the coherent WaveBurst [149, 242], an excess-power algorithm designed for the detection and reconstruction of GWs with minimal assumptions on the waveform morphologies.

In this algorithm, a time-series data is whitened and converted to a time-frequency domain using the Wilson-Daubechies-Meyer wavelet transform [243] with several different time-frequency resolutions. Pixel energy is maximized over all sky positions, and pixels exceeding a specified threshold are retained for each time-frequency map. Subsequently, clustering of neighboring pixels is performed for coincident pixels in the multi-resolution time-frequency maps, and coherent clusters across multiple detectors serve as triggers. Each trigger undergoes evaluation using the constrained maximum likelihood method [244]. For accepted events, waveforms can be reconstructed by applying the inverse wavelet transform to the selected pixels. We use the internal parameters and thresholds employed in the LIGO-Virgo targeted search for CCSNe during the O1 and O2 runs [245]. Figure 7.3 shows the selected pixels and the reconstructed signal of the `he3.5` GW waveform at a distance of 17 kpc, injected into the Einstein Telescope noise.



(a) STFT of noise-added signal

(b) Noise-added and original signals

(c) Selected pixels

(d) Reconstructed signal

FIG. 7.3. Example of reconstruction by the coherent WaveBurst. The original signal is placed at 17 kpc and the detector noise of the Einstein Telescope is added to the signal.

7.3.2 Hilbert-Huang transform

We employ the complementary ensemble empirical mode decomposition [175] to decompose signals into intrinsic mode functions. The decomposition process of a signal $x(t)$ is as follows.

1. Generate N time-series sequences $\{n_k(t)\}$ $(k = 1, \ldots, N)$, where each sequence represents white Gaussian noise with mean zero and standard deviation σ.

2. Construct noise-added data $s_k^{(+)}(t)$ and noise-subtracted data $s_k^{(-)}(t)$ to the target data $s(t)$ for $k = 1, \ldots, N$:

$$s_k^{(\pm)}(t) = s(t) \pm n_k(t). \tag{7.1}$$

3. Apply empirical mode decomposition [56] to each $s_k^{(\pm)}(t)$ to obtain intrinsic mode functions $\{c_{i,k}^{(\pm)}(t)\}$, along with the residue $r_k^{(\pm)}(t)$:

$$s_k^{(\pm)}(t) = \sum_i c_{i,k}^{(\pm)}(t) + r_k^{(\pm)}(t). \tag{7.2}$$

4. Determine the definitive intrinsic mode functions and residue by

$$c_i(t) = \frac{1}{2N} \sum_{k=1}^{N} \Big(c_{i,k}^{(+)}(t) + c_{i,k}^{(-)}(t) \Big), \quad r(t) = \frac{1}{2N} \sum_{k=1}^{N} \Big(r_k^{(+)}(t) + r_k^{(-)}(t) \Big). \tag{7.3}$$

The specific parameters used in this analysis are listed in Table 7.1. After extracting intrinsic mode functions, the instantaneous amplitude $a_i(t)$ and instantaneous frequency $f_i(t)$ of each $c_i(t)$ are computed by the formulae

$$a_i(t) = \sqrt{c_i(t)^2 + \mathcal{H}[c_i(t)]^2}, \tag{7.4}$$

$$f_i(t) = \frac{1}{2\pi} \frac{\mathrm{d}}{\mathrm{d}t} \tan^{-1} \left(\frac{\mathcal{H}[c_i(t)]}{c_i(t)} \right), \tag{7.5}$$

where $\mathcal{H}[\cdot]$ denotes the Hilbert transform,

$$\mathcal{H}[c_i(t)] = \frac{1}{\pi} \mathrm{PV} \int_{-\infty}^{\infty} \mathrm{d}t' \, \frac{c_i(t')}{t - t'}. \tag{7.6}$$

TABLE. 7.1. Parameters used for the complementary ensemble empirical mode decomposition.

Number of Gaussian noise samples	$N = 1000$
Standard deviation of Gaussian noise	$\sigma = 0.8$
Stoppage criterion	$\epsilon = 7 \times 10^{-4}$

7.3.3 Solving universal relation

To infer the physical properties of proto-neutron stars, we employ the universal relation for the ${}^2g_2$ mode [234, 235]:

$$f(t) = 5.88 \times 10^5 x(t) - 86.2 \times 10^6 x^2(t) + 4.67 \times 10^9 x^3(t), \tag{7.7}$$

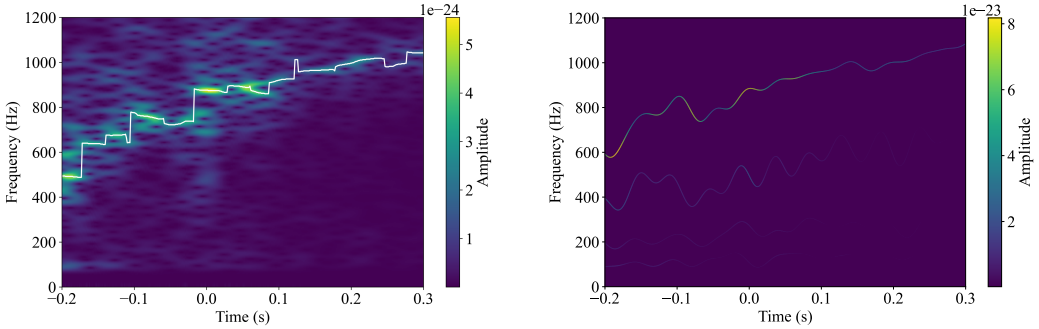
where $f(t)$ is the frequency of the ${}^2g_2$-mode, and $x(t)$ represents the proto-neutron star mass $M(t)$ in solar masses divided by the square of the radius $R(t)$ in kilometers. To determine $x(t)$, we solve Eq. (7.7) for each time using the obtained instantaneous frequency. The solutions to the cubic equations are computed via the eigenvalues of the companion matrix (refer to Appendix F). Finally, we take the average of the estimated values from three detectors of the Einstein Telescope.

7.4 Results

We first present our analyzed results for the raw `he3.5` GW signal. Subsequently, we show the results for the GW signal injected in the Einstein Telescope noise. In the analysis, we generate 100 different Gaussian noise sequences for each distance.

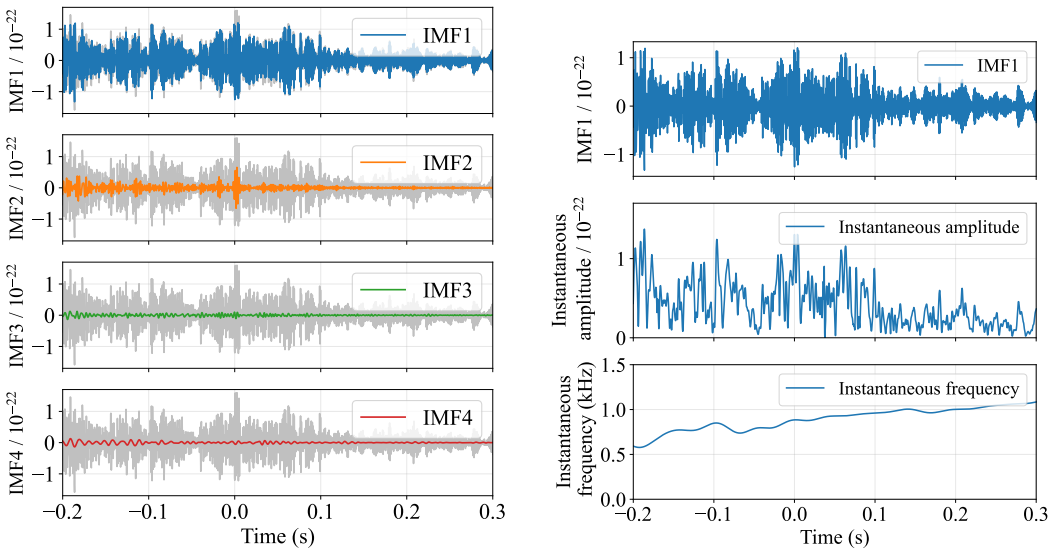
7.4.1 Raw signal

We first applied our approach to the raw signal sample. As the sample does not contain detector noise, we did not apply the coherent WaveBurst reconstruction in this case. Figure 7.4 is the obtained time-frequency map from the Hilbert-Huang transform for the raw `he3.5` signal at a distance of 8 kpc. Additionally, we included the time-frequency map generated using STFT for comparison. In the STFT time-frequency map, a white line corresponds to the frequency with the maximum amplitude for each time bin. This white line and the instantaneous frequency of the first mode in the Hilbert-Huang transform exibit a similar frequency evolution, characterized by a rise from $\sim$ 500 Hz up to $\sim$ 1000 Hz. The extracted intrinsic mode functions, as well as the instantaneous amplitude and frequency of the first mode are plotted in Fig. 7.5. Since the instantaneous amplitudes of the other modes are small compared to that of the first mode, we focus only on the first mode.



(a) Short-time Fourier transform (b) Hilbert-Huang transform

FIG. 7.4. Time-frequency maps of the raw `he3.5` signal from a distance of 8 kpc. The white line in (a) represents the frequency with the maximum amplitude for each time bin.



(a) (b)

FIG. 7.5. Results of the Hilbert-Huang transform of the raw `he3.5` signal. (a) The first four intrinsic mode functions. Each mode is overlaid on the original signal. (b) The instantaneous amplitude and frequency of the first mode.

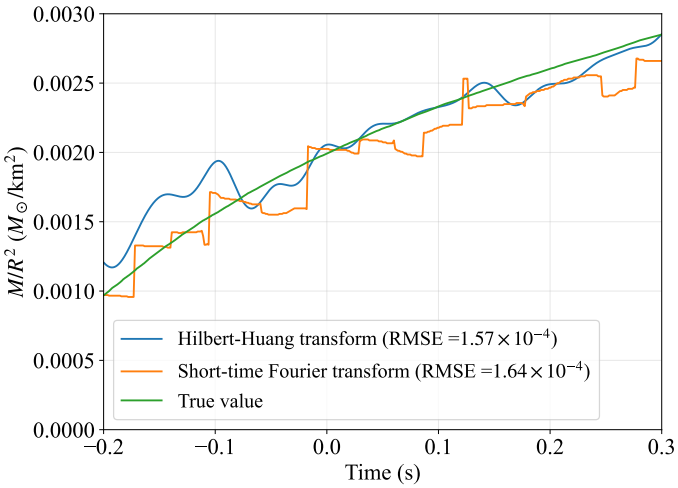


FIG. 7.6. Estimates of M/R^2 of the proto-neutron star from the raw `he3.5` signal.

From these frequencies, we solved the universal relation in Eq. (7.7) for each time. Figure 7.6 shows the computed M/R^2 for each of the extracted frequency from the maximum amplitude in the STFT map, and the instantaneous frequency of the first mode in the Hilbert-Huang transform. The true values are obtained from the numerical simulation. The estimates from the STFT map display some irregularities, yet the values closely approximate the true ones across all time. The estimates from the Hilbert-Huang transform are larger than the true values up to 0.1 s, but beyond that point, they are closer to the true values compared to those from STFT. We also compared the two estimates using the

root mean squared error (RMSE), defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_i (y_i - \hat{y}_i)^2}, \tag{7.8}$$

where N is the total number of samples, y_i represents the true values, and $\hat{y}_i$ is the estimated values. The ratio of the RMSE for the Hilbert-Huang transform to that of STFT was 0.956. This implies that the Hilbert-Huang transform exhibits approximately 4.4% lower error compared to STFT, based on the RMSE values.

7.4.2 Signal in detector noise

For signals injected into detector noise, we first measure how well the reconstructed waveforms match the original waveform. For reconstructed waveform $\{w_k\}$ and injected waveform $\{h_k\}$, the *network overlap* is defined by [246]

$$O_{\mathrm{net}} = \frac{\sum_{k=1}^{K}(w_k|h_k)}{\sqrt{\sum_{k=1}^{K}(w_k|w_k)}\sqrt{\sum_{k=1}^{K}(h_k|h_k)}} \tag{7.9}$$

where K is the number of detectors, i.e. $K = 3$ in this analysis, and $(\cdot|\cdot)$ denotes a noise-weighted inner product defined using the one-sided power spectral density $S(f)$ of the detector noise:

$$(a|b) = 4\mathrm{Re}\int_0^\infty \mathrm{d}f\ \frac{\tilde{a}(f)\tilde{b}^*(f)}{S(f)}. \tag{7.10}$$

The network overlap takes a value between -1 and 1. If $O_{\mathrm{net}} = 1$, the reconstructed waveforms perfectly match the injected waveforms. A value of $O_{\mathrm{net}} = 0$ indicates no match, while $O_{\mathrm{net}} = -1$ implies perfect anti-correlation. Figure 7.7 shows the distributions of the network overlap for each distance or network signal-to-noise ratio (SNR), $\rho_{\mathrm{net}} = (\sum_{k=1}^{K}(h_k|h_k))^{1/2}$. To plot this, we simulated 100 different detector noises and reconstructed the injected waveform in each noise. The plot exhibits the expected behavior: as the distance increases, the overlap decreases, and the variance increases.

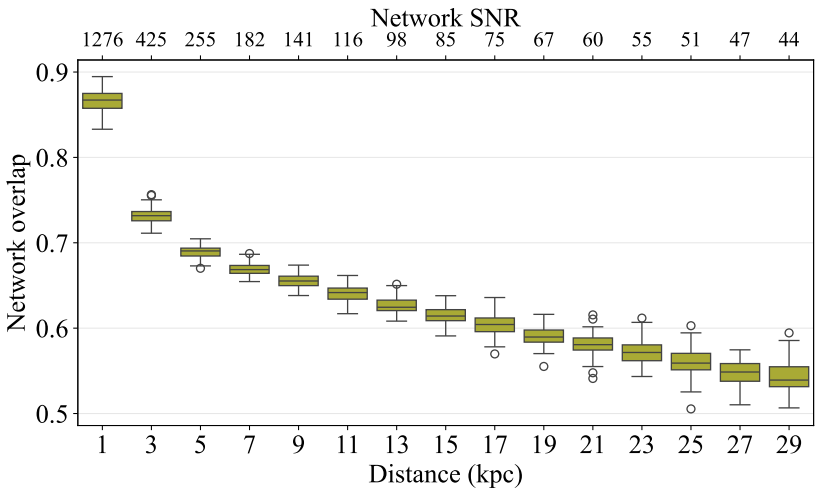


FIG. 7.7. Distributions of the network overlap as a function of distance or network SNR. Each boxplot was generated with 100 different noise realizations.

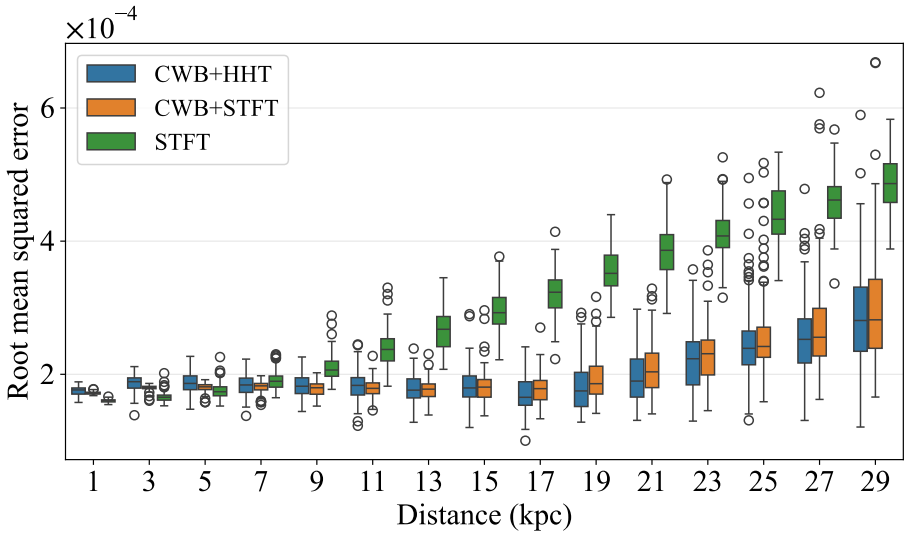


FIG. 7.8. Distributions of the root mean squared error as a function of distance for the three methods. 'CWB+HHT' denotes the use of the coherent WaveBurst reconstruction and the Hilbert-Huang transform. 'CWB+STFT' refers to the use of the coherent WaveBurst reconstruction and the extraction of frequencies from the reconstructed STFT map. 'STFT' denotes the extraction of frequencies from the original STFT map. The white circles in the plot denote the outliers.

We estimated M/R^2 of the proto-neutron stars with three methods. The RMSE for each method is shown in Fig. 7.8. In this plot, the method 'CWB+HHT' denotes the use of the coherent WaveBurst reconstruction and the Hilbert-Huang transform. The second method, 'CWB+STFT', refers to the application of the coherent WaveBurst reconstruction and the extraction of frequencies from the reconstructed STFT map, considering the maximum amplitude for each time bin. The third method, 'STFT' denotes the extraction of frequencies from the original STFT map. For distances less than 5 kpc, the STFT-based method without reconstruction achieves the smallest error. However, as the distance increases, it experiences significantly increased errors, due to the influence of noise other than the g-mode signal. As the distance increases, the two methods employing the coherent WaveBurst reconstruction reduce the effect of noise, through the extraction of coherent g-mode signals across multiple detectors. At shorter distances, the estimates from the Hilbert-Huang transform have larger errors than those from the STFT, but this trend reverses as the distance becomes larger. Figure 7.9 presents the average values of the RMSE for each distance using these two methods. From this figure, the STFT-based method outperforms for distances less than $\sim$ 15 kpc, while at greater distances, HHT surpasses. Based on these results, we conclude that these two methods provide comparable estimations for M/R^2. Additionally, in Fig. 7.10, we plot the RMSE against the network overlap for the estimates from the Hilbert-Huang transform. For samples with a network overlap smaller than around 0.6, the RMSE is significantly larger. From these findings, it is evident that accurate waveform extraction is crucial for estimating M/R^2.

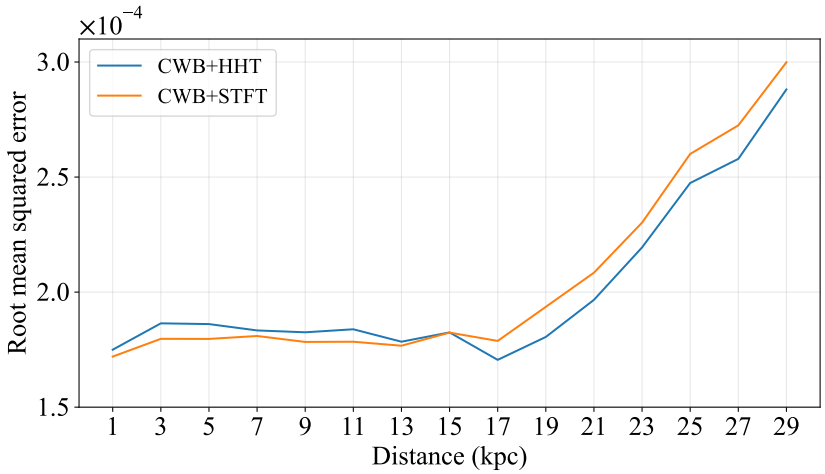


FIG. 7.9. Average of the root mean squared error for each distance.

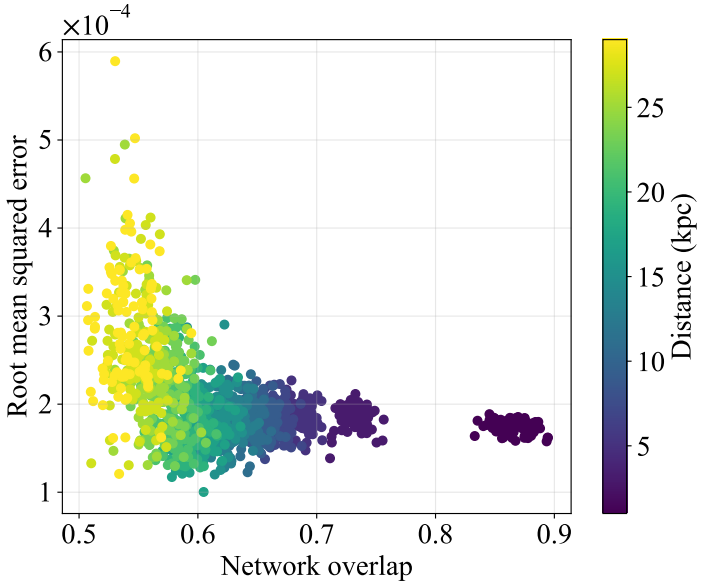


FIG. 7.10. Root mean squared error against the network overlap for the estimates derived from the Hilbert-Huang transform.

7.5 Discussion

The estimates from the raw signal using the Hilbert-Huang transform exhibited an overestimation of the true values during the initial 0.1 s. The presence of the ${}^2g_1$ mode is considered as a potential cause. The universal relation for the ${}^2g_1$ mode is expressed as [234, 235]

$$f(t) = 8.67 \times 10^5 x(t) - 5.19 \times 10^7 x^2(t). \tag{7.11}$$

Using the simulated values of M/R^2 and the universal relations for the ${}^2g_1$ and ${}^2g_2$ modes, we plotted the frequencies of these two modes on the STFT map of the raw signal, as shown in Fig. 7.11. Additionally, Fig. 7.12 displays the computed instantaneous frequency of the first mode in the Hilbert-Huang transform, as well as the frequencies of these two modes. During the time when the frequency of the estimated ${}^2g_1$ mode is within the considered band (100-1200 Hz), the frequency of the first mode in the Hilbert-Huang

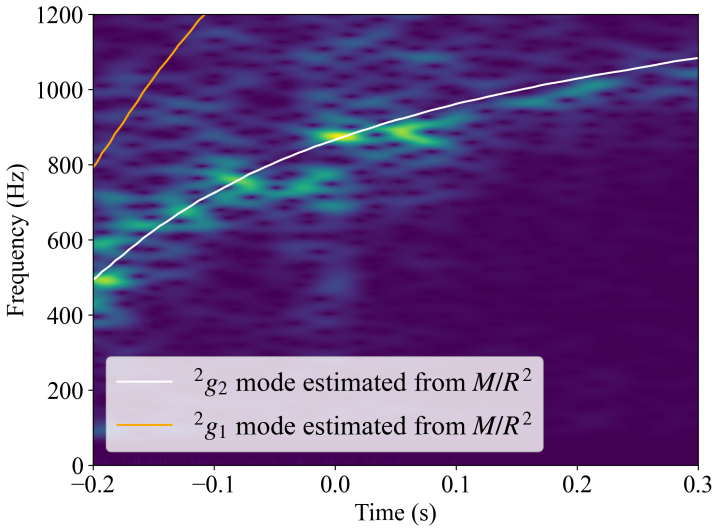


FIG. 7.11. Two modes calculated using the universal relations from the simulated M/R^2. The time-frequency map is derived from the raw *he3.5* signal.

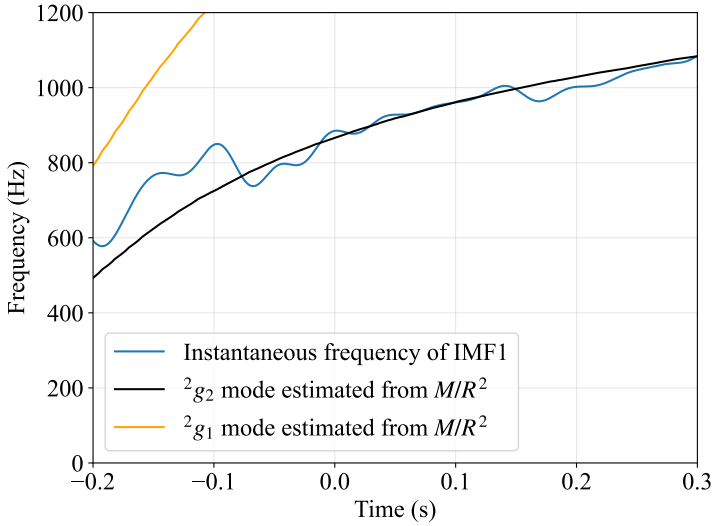


FIG. 7.12. Instantaneous frequency of the first intrinsic mode function along with the $^2g_1$ and $^2g_2$ modes derived from the simulated M/R^2.

transform is higher than that of the $^2g_2$ mode. Therefore, during this time period, the first mode extracted in the Hilbert-Huang transform may be the mixed mode of these two modes. The handling of this mixed mode or potential split mode will be discussed in future research.

7.6 Conclusions

In this study, we demonstrated the use of the Hilbert-Huang transform for estimating M/R^2 of proto-neutron stars from their g-mode GW signals. We first applied it to the raw signal and confirmed the effectiveness of our method to extract the g-mode frequencies. We then tested our approach, the combination of the coherent WaveBurst reconstruction and the Hilbert-Huang transform, for the signal injected into the detector noise at the

Einstein Telescope. Our new approach showed comparable results to the STFT-based one in terms of the RMSE.

While our approach may not have demonstrated significantly better accuracy than STFT, having two independent methods that yield comparable accuracy is valuable in real-world applications where the actual value is unknown. To validate the efficacy of our method further, testing with additional simulation data is necessary. For waveform reconstruction, we utilized the coherent WaveBurst algorithm, which does not assume a waveform model in advance. Additionally, we would like to compare it with a reconstruction method that assumes a specific waveform model, as demonstrated in Ref. [238].

7.7 Appendix: Sample data and results

We randomly selected one sample from each of the 100 datasets at each distance and plotted the estimates of M/R^2 along with the STFT map in Figs. 7.13 to 7.17.

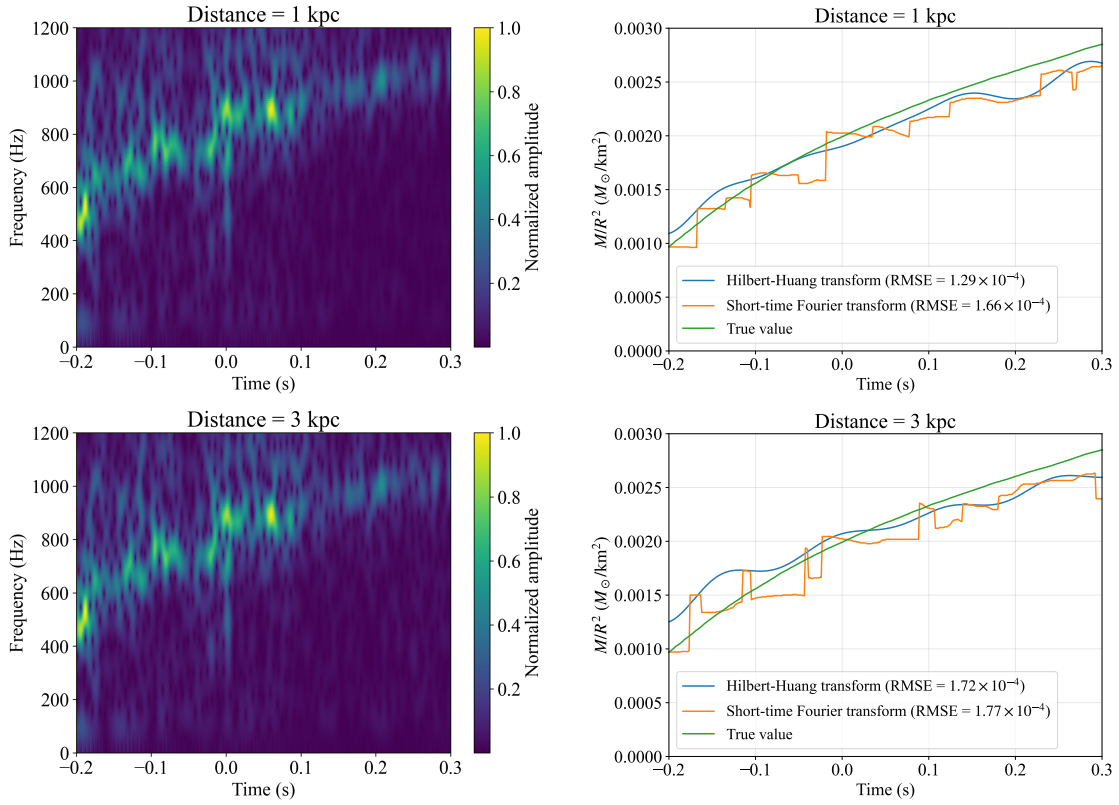


FIG. 7.13. Test samples in STFT maps and the estimates of M/R^2 at distances of 1 and 3 kpc.

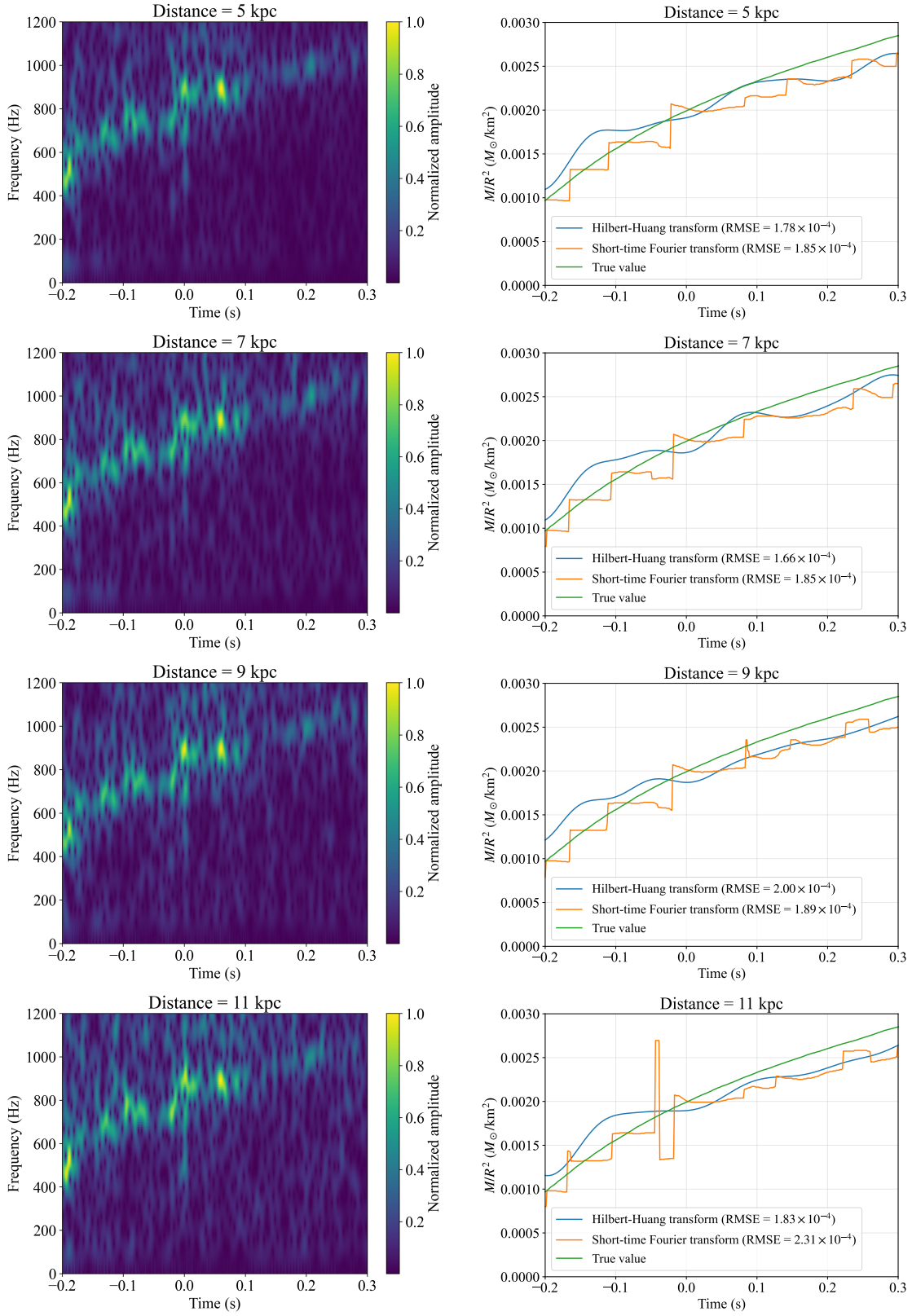


FIG. 7.14. Test samples in STFT maps and the estimates of M/R^2 at distances of 5, 7, 9, and 11 kpc.

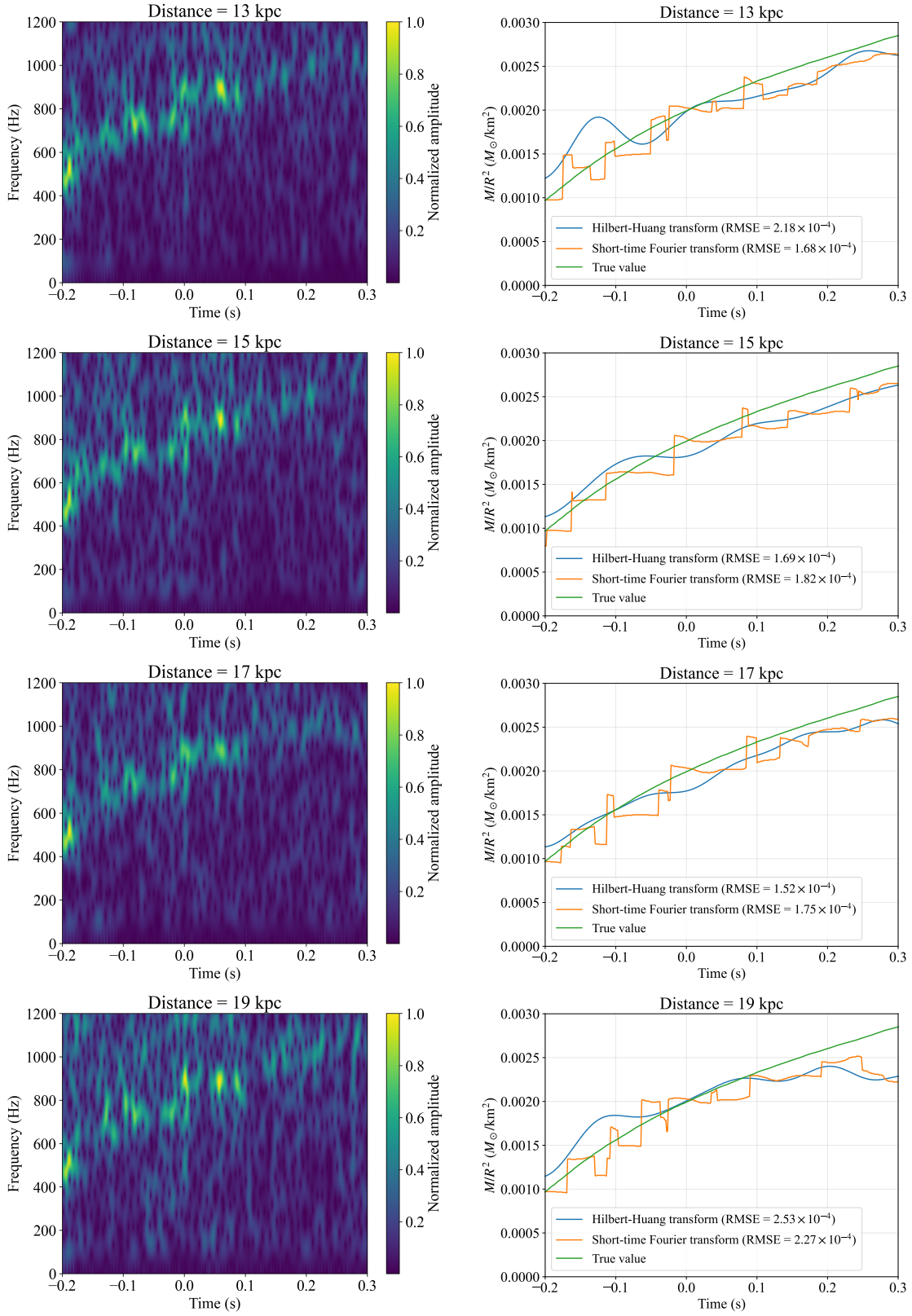


FIG. 7.15. Test samples in STFT maps and the estimates of M/R^2 at distances of 13, 15, 17, and 19 kpc.

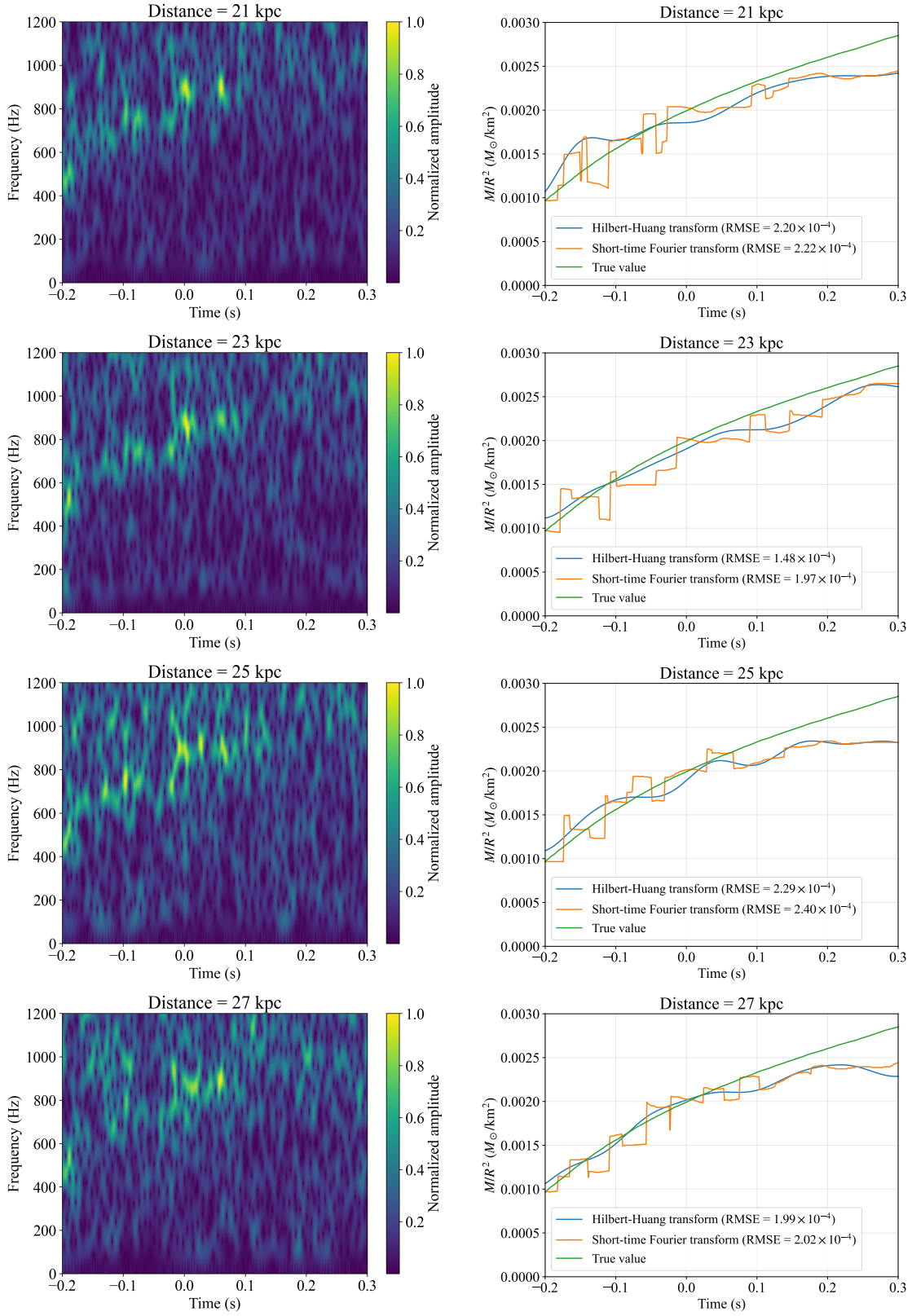


FIG. 7.16. Test samples in STFT maps and the estimates of M/R^2 at distances of 21, 23, 25, and 27 kpc.

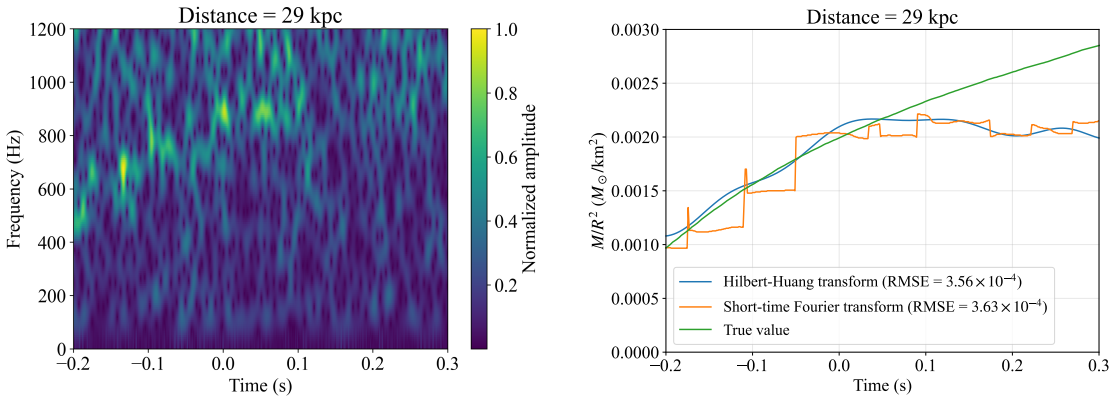


FIG. 7.17. Test samples in STFT maps and the estimates of M/R^2 at a distance of 29 kpc.

Chapter 8

Conclusions

8.1 Summary of research results

This thesis explored the applications of novel techniques, including deep learning and the Hilbert-Huang transform, to the analysis of gravitational-wave (GW) data. In the study presented in Chapter 5, a convolutional neural network (CNN) model was trained using time-frequency maps to classify GW signals from core-collapse supernovae (CCSNe) within the detector noise of the third observing run. The model's interpretability was enhanced through the t-distributed stochastic neighbor embedding and the class activation mapping (CAM) techniques. Notably, this study represents the first application of CAM techniques to GW signal analysis in this field. Visualization via CAM techniques revealed the model's reliance on specific features in the spectrogram, especially the g-mode.

Chapter 6 investigated the distinctions between one-dimensional and two-dimensional CNNs for detecting GWs from compact binary coalescences (CBCs). The 1D model demonstrated superior performance in identifying binary black hole signals, while the 2D model excelled with the other two types of signals. The computations of feature importance revealed the unique strengths of the 1D and 2D models, highlighting the benefits of their integration. The ensemble model, combining the strengths of both individual models, demonstrated robust classification across all signal types.

In Chapter 7, the Hilbert-Huang transform was applied to estimate M/R^2 from the g-mode GW signals, where M and R denote the mass and the radius of proto-neutron stars. The method was validated on raw signals and tested on signals injected into detector noise, showing comparable results to short-time Fourier transform (STFT)-based approach, highlighting its potential as a new method for such parameter estimations.

8.2 Future prospects

Here, we discuss the future prospects for each of the three studies described in this thesis.

8.2.1 Core-collapse supernova detection and classification

As mentioned in Chapter 5, the CAM visualization indicated that certain low-frequency modes were not utilized by the model for predictions. Additional or alternative time-frequency maps from wavelet and the Hilbert-Huang transforms should be considered. If applying the Hilbert-Huang transform, it is necessary to suppress noise or reconstruct

waveforms beforehand, and the methodology for these processes should be carefully considered.

Preparing enough samples for training and validating a model for GWs from CCSN explosions is challenging due to the absence of observations. Furthermore, the availability of multi-dimensional simulation data is limited. To improve the detection model, machine learning techniques can be used to generate new waveforms for the training set. Multiple groups have already studied the use of a generative adversarial network [169] for glitch noise generation [168, 247, 248]. GW signals from CCSNe could be generated in a similar way.

8.2.2 Compact binary coalescence detection and classification

The evaluation of our models outlined in Chapter 6 is not complete. Given that CBC events have been detected over 100 times, we can assess the validity of our method using the observed data. Our classification model can be employed using a sliding window approach for long-duration data. It is important to consider the sliding steps and the treatment of triggers. Additionally, applying the model to background noise is crucial for computing the false alarm rate.

In the field of GW detection, machine learning has two primary objectives: to speed up the identification of signals and to detect signals that may be missed by other pipelines. In the latter scenario, training models focused on complex signals that incorporate physical effects of interest, such as eccentricity, precession, and higher-order modes, may discover events that were previously missed.

8.2.3 Parameter estimation of proto-neutron star

The study discussed in Chapter 7 requires further investigation. First, we need to validate our approach using a wider range of simulated signals. While we utilized the coherent WaveBurst algorithm for waveform reconstruction, exploring alternative methods is essential. For instance, the BayesWave algorithm [249–251] is another method that can reconstruct GW signals from noise and glitches without any prior waveform morphologies. Reconstruction with an assumption on the signal model was performed in Ref. [238] through Bayesian parameter estimation using an asymmetric chirplet signal model [252] and the non-central chi-squared likelihood method [253]. Comparing these reconstruction methods before applying the Hilbert-Huang transform can provide valuable insights. Although we employed the Einstein Telescope detectors, exploring other detectors, including current ones and the Cosmic Explorer, adds an interesting dimension to the analysis.

In Chapter 7, we mentioned the potential occurrence of mode mixing in the first intrinsic mode function, which requires further consideration. If the ${}^2g_1$ mode is indeed present in the GW signal we used, it may be possible, by tuning the parameters of the ensemble empirical mode decomposition, to extract an intrinsic mode function that expresses the ${}^2g_1$ mode for the initial 0.1 s and then transitions to the ${}^2g_2$ mode. In such a scenario, the first 0.1 s of the ${}^2g_2$ mode would appear in the second intrinsic mode function, which means that the ${}^2g_2$ mode is split in two intrinsic mode functions. The challenge lies in how to connect the two split modes in this case.

Moreover, it is necessary to investigate whether our method can extract multiple modes when using other simulation data in which multiple modes appear. The method we used for comparison, which extracts the frequencies from the maximum amplitude at each time bin of the STFT image, cannot extract multiple modes; HHT is considered advantageous in this respect. Furthermore, it would be beneficial to explore whether the estimation of M/R^2 using multiple modes improves the accuracy compared to using a single mode.

Appendix A

Expansion of Ricci Tensor

Here, we compute the expansion of the Ricci tensor up to the second order around the background spacetime, following the procedure outlined in Ref. [254]. The results are used in expanding the Einstein-Hilbert action, as shown in Eq. (2.8).

We write the metric tensor $g_{\mu\nu}$ as the sum of the background $g^{(\mathrm{B})}_{\mu\nu}$ and the perturbative term $h_{\mu\nu}$,

$$g_{\mu\nu} = g^{(\mathrm{B})}_{\mu\nu} + h_{\mu\nu}. \tag{A.1}$$

We expand the Ricci tensor and the Christoffel symbol in terms of the $h_{\mu\nu}$ as

$$R_{\mu\nu} = R^{(\mathrm{B})}_{\mu\nu} + R^{(1)}_{\mu\nu} + R^{(2)}_{\mu\nu} + \mathcal{O}(h^3), \tag{A.2}$$

$$\Gamma^{\alpha}{}_{\mu\nu} = \Gamma^{(\mathrm{B})\alpha}{}_{\mu\nu} + \Gamma^{(1)\alpha}{}_{\mu\nu} + \Gamma^{(2)\alpha}{}_{\mu\nu} + \mathcal{O}(h^3). \tag{A.3}$$

The connection $\Gamma^{\alpha}{}_{\mu\nu}$ and the Riemann tensor $R_{\mu\nu}$ are defined for the metric $g_{\mu\nu}$, while $\Gamma^{(\mathrm{B})\alpha}{}_{\mu\nu}$ and $R^{(\mathrm{B})}_{\mu\nu}$ are defined for the background metric $g^{(\mathrm{B})}_{\mu\nu}$. Covariant derivatives with respect to the metric $g_{\mu\nu}$ is denoted by ';', and covariant derivative with respect to the metric $g^{(\mathrm{B})}_{\mu\nu}$ is denoted by '—'. For instance,

$$A^{\mu}{}_{;\alpha} = A^{\mu}{}_{,\alpha} + A^{\nu}\Gamma^{\mu}{}_{\nu\alpha}, \tag{A.4}$$

$$A^{\mu}{}_{|\alpha} = A^{\mu}{}_{,\alpha} + A^{\nu}\Gamma^{(\mathrm{B})\mu}{}_{\nu\alpha}. \tag{A.5}$$

The raising and lowering of indices for $h_{\mu\nu}$ are performed using $g^{(B)\mu\nu}$ as follows,

$$g^{(\mathrm{B})\mu\nu}h_{\nu\rho} = h^{\mu}{}_{\rho}. \tag{A.6}$$

We first note that $g^{\mu\nu}$ is expanded as

$$g^{\mu\nu} = g^{(\mathrm{B})\mu\nu} - h^{\mu\nu} + h^{\mu\alpha}h_{\alpha}{}^{\nu} - h^{\mu\alpha}h_{\alpha}{}^{\beta}h_{\beta}{}^{\nu} + \mathcal{O}(h^4). \tag{A.7}$$

This expansion can be confirmed by proving $g_{\rho\mu}g^{\mu\nu} = \delta_{\rho}{}^{\nu}$. In practice,

$$\begin{aligned} g_{\rho\mu}g^{\mu\nu} &= (g^{(\mathrm{B})}_{\rho\mu} + h_{\rho\mu})(g^{(\mathrm{B})\mu\nu} - h^{\mu\nu} + h^{\mu\alpha}h_{\alpha}{}^{\nu} - h^{\mu\alpha}h_{\alpha}{}^{\beta}h_{\beta}{}^{\nu}) \\ &= (\delta_{\rho}{}^{\nu} - h_{\rho}{}^{\nu} + h_{\rho}{}^{\alpha}h_{\alpha}{}^{\nu} - h_{\rho}{}^{\alpha}h_{\alpha}{}^{\beta}h_{\beta}{}^{\nu}) + (h_{\rho}{}^{\nu} - h_{\rho\mu}h^{\mu\nu} + h_{\rho\mu}h^{\mu\alpha}h_{\alpha}{}^{\nu}) \\ &= \delta_{\rho}{}^{\nu} + \mathcal{O}(h^4). \end{aligned} \tag{A.8}$$

We define $S^{\mu}{}_{\beta\gamma}$ as the difference between the Christoffel symbol $\Gamma^{\mu}{}_{\beta\gamma}$ and the background Christoffel symbol $\Gamma^{(\mathrm{B})\mu}{}_{\beta\gamma}$:

$$S^{\mu}{}_{\beta\gamma} = \Gamma^{\mu}{}_{\beta\gamma} - \Gamma^{(\mathrm{B})\mu}{}_{\beta\gamma}. \tag{A.9}$$

From the transformation law for the Christoffel symbol,

$$\Gamma^{\mu'}{}_{\beta'\gamma'} = \frac{\partial x^{\mu'}}{\partial x^{\mu}} \frac{\partial x^{\beta}}{\partial x^{\beta'}} \frac{\partial x^{\gamma}}{\partial x^{\gamma'}} \Gamma^{\mu}{}_{\beta\gamma} + \frac{\partial x^{\mu'}}{\partial x^{\alpha}} \frac{\partial^2 x^{\alpha}}{\partial x^{\beta'} \partial x^{\gamma'}}, \tag{A.10}$$

$$\Gamma^{(\mathrm{B})\mu'}{}_{\beta'\gamma'} = \frac{\partial x^{\mu'}}{\partial x^{\mu}} \frac{\partial x^{\beta}}{\partial x^{\beta'}} \frac{\partial x^{\gamma}}{\partial x^{\gamma'}} \Gamma^{(\mathrm{B})\mu}{}_{\beta\gamma} + \frac{\partial x^{\mu'}}{\partial x^{\alpha}} \frac{\partial^2 x^{\alpha}}{\partial x^{\beta'} \partial x^{\gamma'}}, \tag{A.11}$$

the transformation of $S^{\mu}{}_{\beta\gamma}$ becomes

$$S^{\mu'}{}_{\beta'\gamma'} = \frac{\partial x^{\mu'}}{\partial x^{\mu}} \frac{\partial x^{\beta}}{\partial x^{\beta'}} \frac{\partial x^{\gamma}}{\partial x^{\gamma'}} S^{\mu}{}_{\beta\gamma}. \tag{A.12}$$

Thus, $S^{\mu}{}_{\beta\gamma}$ is a tensor, and $S_{\alpha\beta\gamma}$ is calculated as

$$\begin{aligned} S_{\alpha\beta\gamma} &= g_{\alpha\nu} S^{\nu}{}_{\beta\gamma} \\ &= g_{\alpha\nu} \Gamma^{\nu}{}_{\beta\gamma} - (g^{(\mathrm{B})}_{\alpha\nu} + h_{\alpha\nu}) \Gamma^{(\mathrm{B})\nu}{}_{\beta\gamma} \\ &= \frac{1}{2}(h_{\alpha\beta,\gamma} + h_{\alpha\gamma,\beta} - h_{\beta\gamma,\alpha}) - h_{\alpha\nu} \Gamma^{(\mathrm{B})\nu}{}_{\beta\gamma}. \end{aligned} \tag{A.13}$$

Using the identity

$$\begin{aligned} h_{\alpha\beta|\gamma} + h_{\alpha\gamma|\beta} - h_{\beta\gamma|\alpha} &= h_{\alpha\beta,\gamma} - h_{\nu\beta} \Gamma^{(\mathrm{B})\nu}{}_{\alpha\gamma} - h_{\alpha\nu} \Gamma^{(\mathrm{B})\nu}{}_{\beta\gamma} \\ &\quad + h_{\alpha\gamma,\beta} - h_{\nu\gamma} \Gamma^{(\mathrm{B})\nu}{}_{\alpha\beta} - h_{\alpha\nu} \Gamma^{(\mathrm{B})\nu}{}_{\gamma\beta} \\ &\quad - h_{\beta\gamma,\alpha} + h_{\nu\gamma} \Gamma^{(\mathrm{B})\nu}{}_{\beta\alpha} + h_{\beta\nu} \Gamma^{(\mathrm{B})\nu}{}_{\gamma\alpha} \\ &= h_{\alpha\beta,\gamma} + h_{\alpha\gamma,\beta} - h_{\beta\gamma,\alpha} - 2h_{\alpha\nu} \Gamma^{(\mathrm{B})\nu}{}_{\beta\gamma}, \end{aligned} \tag{A.14}$$

we obtain the following:

$$S_{\alpha\beta\gamma} = \frac{1}{2}(h_{\alpha\beta|\gamma} + h_{\alpha\gamma|\beta} - h_{\beta\gamma|\alpha}), \tag{A.15}$$

$$S^{\mu}{}_{\beta\gamma} = \frac{1}{2} g^{\mu\alpha}(h_{\alpha\beta|\gamma} + h_{\alpha\gamma|\beta} - h_{\beta\gamma|\alpha}). \tag{A.16}$$

Next, we proceed to compute the Riemann tensor. The Riemann tensor $R^{\alpha}{}_{\beta\gamma\delta}$ for the metric $g_{\mu\nu}$ and the background Riemann tensor $R^{(\mathrm{B})\alpha}{}_{\beta\gamma\delta}$ for the background metric $g^{(\mathrm{B})}_{\mu\nu}$ are defined as

$$R^{\alpha}{}_{\beta\gamma\delta} = \Gamma^{\alpha}{}_{\beta\delta,\gamma} - \Gamma^{\alpha}{}_{\beta\gamma,\delta} + \Gamma^{\alpha}{}_{\mu\gamma} \Gamma^{\mu}{}_{\beta\delta} - \Gamma^{\alpha}{}_{\mu\delta} \Gamma^{\mu}{}_{\beta\gamma}, \tag{A.17}$$

$$R^{(\mathrm{B})\alpha}{}_{\beta\gamma\delta} = \Gamma^{(\mathrm{B})\alpha}{}_{\beta\delta,\gamma} - \Gamma^{(\mathrm{B})\alpha}{}_{\beta\gamma,\delta} + \Gamma^{(\mathrm{B})\alpha}{}_{\mu\gamma} \Gamma^{(\mathrm{B})\mu}{}_{\beta\delta} - \Gamma^{(\mathrm{B})\alpha}{}_{\mu\delta} \Gamma^{(\mathrm{B})\mu}{}_{\beta\gamma}. \tag{A.18}$$

The purtabative term of the Riemann tensor can be expressed using S as

$$\begin{aligned}
R^{\alpha}{}_{\beta\gamma\delta} - R^{(\mathrm{B})\alpha}{}_{\beta\gamma\delta} &= (\Gamma^{\alpha}{}_{\beta\delta,\gamma} - \Gamma^{(\mathrm{B})\alpha}{}_{\beta\delta,\gamma}) - (\Gamma^{\alpha}{}_{\beta\gamma,\delta} - \Gamma^{(\mathrm{B})\alpha}{}_{\beta\gamma,\delta}) \\
&\quad + \Gamma^{\alpha}{}_{\mu\gamma}\Gamma^{\mu}{}_{\beta\delta} - \Gamma^{\alpha}{}_{\mu\delta}\Gamma^{\mu}{}_{\beta\gamma} - (\Gamma^{(\mathrm{B})\alpha}{}_{\mu\gamma}\Gamma^{(\mathrm{B})\mu}{}_{\beta\delta} - \Gamma^{(\mathrm{B})\alpha}{}_{\mu\delta}\Gamma^{(\mathrm{B})\mu}{}_{\beta\gamma}) \\
&= S^{\alpha}{}_{\beta\delta,\gamma} - S^{\alpha}{}_{\beta\gamma,\delta} + S^{\alpha}{}_{\mu\gamma}S^{\mu}{}_{\beta\delta} - S^{\alpha}{}_{\mu\delta}S^{\mu}{}_{\beta\gamma} \\
&\quad + \Gamma^{\alpha}{}_{\mu\gamma}\Gamma^{(\mathrm{B})\mu}{}_{\beta\delta} + \Gamma^{\mu}{}_{\beta\delta}\Gamma^{(\mathrm{B})\alpha}{}_{\mu\gamma} - 2\Gamma^{(\mathrm{B})\alpha}{}_{\mu\gamma}\Gamma^{(\mathrm{B})\mu}{}_{\beta\delta} \\
&\quad - \Gamma^{\alpha}{}_{\mu\delta}\Gamma^{(\mathrm{B})\mu}{}_{\beta\gamma} - \Gamma^{\mu}{}_{\beta\gamma}\Gamma^{(\mathrm{B})\alpha}{}_{\mu\delta} + 2\Gamma^{(\mathrm{B})\alpha}{}_{\mu\delta}\Gamma^{(\mathrm{B})\mu}{}_{\beta\gamma} \qquad \text{(A.19)} \\
&= S^{\alpha}{}_{\beta\delta|\gamma} - S^{\alpha}{}_{\beta\gamma|\delta} + S^{\alpha}{}_{\mu\gamma}S^{\mu}{}_{\beta\delta} - S^{\alpha}{}_{\mu\delta}S^{\mu}{}_{\beta\gamma}. \qquad \text{(A.20)}
\end{aligned}$$

Thus, the Ricci tensor is expressed as

$$R_{\mu\nu} - R^{(\mathrm{B})}_{\mu\nu} = S^{\alpha}{}_{\mu\nu|\alpha} - S^{\alpha}{}_{\mu\alpha|\nu} + S^{\alpha}{}_{\beta\alpha}S^{\beta}{}_{\mu\nu} - S^{\alpha}{}_{\beta\nu}S^{\beta}{}_{\mu\alpha}. \qquad \text{(A.21)}$$

We compute each term in the right-hand side of this equation.

$$\begin{aligned}
S^{\alpha}{}_{\mu\nu|\alpha} &= \frac{1}{2}\{g^{\alpha\beta}(h_{\beta\mu|\nu} + h_{\beta\nu|\mu} - h_{\mu\nu|\beta})\}_{|\alpha} \\
&= \frac{1}{2}\{(g^{(\mathrm{B})\alpha\beta} - h^{\alpha\beta})(h_{\beta\mu|\nu} + h_{\beta\nu|\mu} - h_{\mu\nu|\beta})\}_{|\alpha} \\
&= \frac{1}{2}(h^{\alpha}{}_{\mu|\nu} + h^{\alpha}{}_{\nu|\mu} - h_{\mu\nu}{}^{|\alpha} - h^{\alpha\beta}h_{\beta\mu|\nu} - h^{\alpha\beta}h_{\beta\nu|\mu} + h^{\alpha\beta}h_{\mu\nu|\beta})_{|\alpha} \qquad \text{(A.22)} \\
S^{\alpha}{}_{\mu\alpha|\nu} &= \frac{1}{2}(h^{\alpha}{}_{\mu|\alpha} + h^{\alpha}{}_{\alpha|\mu} - h_{\mu\alpha}{}^{|\alpha} - h^{\alpha\beta}h_{\beta\mu|\alpha} - h^{\alpha\beta}h_{\beta\alpha|\mu} + h^{\alpha\beta}h_{\mu\alpha|\beta})_{|\nu} \qquad \text{(A.23)}
\end{aligned}$$

Since S is proportional to h, the linear term in h of the Ricci tensor comes from the first two terms of the right-hand side of Eq. (A.21). We obtain

$$\begin{aligned}
R^{(1)}_{\mu\nu} &= S^{(1)\alpha}{}_{\mu\nu|\alpha} - S^{(1)\alpha}{}_{\mu\alpha|\nu} \\
&= \frac{1}{2}(-h_{|\mu\nu} - h_{\mu\nu|\alpha}{}^{\alpha} + h_{\alpha\mu|\nu}{}^{\alpha} + h_{\alpha\nu|\mu}{}^{\alpha}). \qquad \text{(A.24)}
\end{aligned}$$

Next, we compute $R^{(2)}_{\mu\nu}$. The contributions from the first and second terms in Eq. (A.21) are given by

$$\begin{aligned}
&S^{(2)\alpha}{}_{\mu\nu|\alpha} - S^{(2)\alpha}{}_{\mu\alpha|\nu} \\
&= -\frac{1}{2}h^{\alpha\beta}{}_{|\alpha}(h_{\beta\mu|\nu} + h_{\beta\nu|\mu} - h_{\mu\nu|\beta}) - \frac{1}{2}h^{\alpha\beta}(h_{\beta\mu|\nu\alpha} + h_{\beta\nu|\mu\alpha} - h_{\mu\nu|\beta\alpha}) \\
&\quad + \frac{1}{2}h^{\alpha\beta}{}_{|\nu}(h_{\beta\mu|\alpha} + h_{\beta\alpha|\mu} - h_{\mu\alpha|\beta}) + \frac{1}{2}h^{\alpha\beta}(h_{\beta\mu|\alpha\nu} + h_{\beta\alpha|\mu\nu} - h_{\mu\alpha|\beta\nu}) \\
&= \frac{1}{2}h^{\alpha\beta}(h_{\alpha\beta|\mu\nu} + h_{\mu\nu|\alpha\beta} - h_{\alpha\mu|\nu\beta} - h_{\alpha\nu|\mu\beta}) + \frac{1}{2}h_{\alpha\beta|\mu}h^{\alpha\beta}{}_{|\nu} \\
&\quad - \frac{1}{2}h^{\alpha\beta}{}_{|\beta}(h_{\alpha\mu|\nu} + h_{\alpha\nu|\mu} - h_{\mu\nu|\alpha}). \qquad \text{(A.25)}
\end{aligned}$$

The third and fourth terms in Eq. (A.21) are computed as

$$\begin{aligned} S^{\alpha}{}_{\beta\alpha}S^{\beta}{}_{\mu\nu} &= \frac{1}{2}g^{\alpha\rho}(h_{\rho\beta|\alpha} + h_{\rho\alpha|\beta} - h_{\beta\alpha|\rho})\frac{1}{2}g^{\beta\rho}(h_{\rho\mu|\nu} + h_{\rho\nu|\mu} - h_{\mu\nu|\rho}) \\ &= \frac{1}{4}(h^{\alpha}{}_{\beta|\alpha} + h^{\alpha}{}_{\alpha|\beta} - h_{\beta\alpha}{}^{|\alpha})(h^{\beta}{}_{\mu|\nu} + h^{\beta}{}_{\nu|\mu} - h_{\mu\nu}{}^{|\beta}) + \mathcal{O}(h^3) \qquad \text{(A.26)} \\ S^{\alpha}{}_{\beta\nu}S^{\beta}{}_{\mu\alpha} &= \frac{1}{4}(h^{\alpha}{}_{\beta|\nu} + h^{\alpha}{}_{\nu|\beta} - h_{\beta\nu}{}^{|\alpha})(h^{\beta}{}_{\mu|\alpha} + h^{\beta}{}_{\alpha|\mu} - h_{\mu\alpha}{}^{|\beta}) + \mathcal{O}(h^3). \qquad \text{(A.27)} \end{aligned}$$

Thus,

$$\begin{aligned} &S^{(1)\alpha}{}_{\beta\alpha}S^{(1)\beta}{}_{\mu\nu} - S^{(1)\alpha}{}_{\beta\nu}S^{(1)\beta}{}_{\mu\alpha} \\ &= \frac{1}{4}h_{|\beta}(h^{\beta}{}_{\mu|\nu} + h^{\beta}{}_{\nu|\mu} - h_{\mu\nu}{}^{|\beta}) - \frac{1}{4}h_{\alpha\beta|\mu}h^{\alpha\beta}{}_{|\nu} - \frac{1}{2}h_{\nu}{}^{\alpha|\beta}(h_{\beta\mu|\alpha} - h_{\alpha\mu|\beta}) \qquad \text{(A.28)} \end{aligned}$$

Therefore, the final result for $R^{(2)}_{\mu\nu}$ is given by

$$\begin{aligned} R^{(2)}_{\mu\nu} = \frac{1}{2}\Bigg[&\frac{1}{2}h_{\alpha\beta|\mu}h^{\alpha\beta}{}_{|\nu} + h^{\alpha\beta}(h_{\alpha\beta|\mu\nu} + h_{\mu\nu|\alpha\beta} - h_{\alpha\mu|\nu\beta} - h_{\alpha\nu|\mu\beta}) \\ &+h_{\nu}{}^{\alpha|\beta}(h_{\alpha\mu|\beta} - h_{\beta\mu|\alpha}) - \left(h^{\alpha\beta}{}_{|\beta} - \frac{1}{2}h^{|\alpha}\right)(h_{\alpha\mu|\nu} + h_{\alpha\nu|\mu} - h_{\mu\nu|\alpha})\Bigg]. \qquad \text{(A.29)} \end{aligned}$$

Appendix B

Gravitational-Wave Calculation in Arbitrary Directions

Here, we derive the explicit forms of the plus and cross polarizations of GW signals radiated in arbitrary directions. The resulting equations are important in calculating GW signals from the quadrupole moments of CCSNe obtained through 3D numerical simulations.

The quadrupole formula of GWs is given by

$$h_{ij}(\boldsymbol{x},t) = \frac{2G}{c^4 r}\ddot{Q}_{ij}\Big(t-\frac{r}{c}\Big). \tag{B.1}$$

Let us start by deriving the amplitude of a GW propagating in z direction. The matrix P_{ij} defined by Eq. (2.36) is then

$$P_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{B.2}$$

We can transform the quadrupole moments into the transverse traceless gauge using the Lambda tensor defined in Eq. (2.35) as

$$\begin{aligned} \Lambda_{ij,kl}\ddot{Q}_{kl} &= \Big(P_{ik}P_{jl} - \frac{1}{2}P_{ij}P_{kl}\Big)\ddot{Q}_{kl} \\ &= (P\ddot{Q}P)_{ij} - \frac{1}{2}P_{ij}(\ddot{Q}_{xx}+\ddot{Q}_{yy}) \\ &= \begin{pmatrix} (\ddot{Q}_{xx}-\ddot{Q}_{yy})/2 & \ddot{Q}_{xy} & 0 \\ \ddot{Q}_{xy} & -(\ddot{Q}_{xx}-\ddot{Q}_{yy})/2 & 0 \\ 0 & 0 & 0 \end{pmatrix}_{ij} \end{aligned} \tag{B.3}$$

The plus and cross polarizations are then defined by

$$h_+ = \frac{G}{c^4 r}(\ddot{Q}_{xx}-\ddot{Q}_{yy}), \tag{B.4}$$

$$h_\times = \frac{2G}{c^4 r}\ddot{Q}_{xy}. \tag{B.5}$$

Here, the right-hand side is evaluated at a time $t - r/c$. Next, let us compute the GW signal propagating in an arbitrary direction n. We define θ and ϕ as illustrated in Fig. B.1.

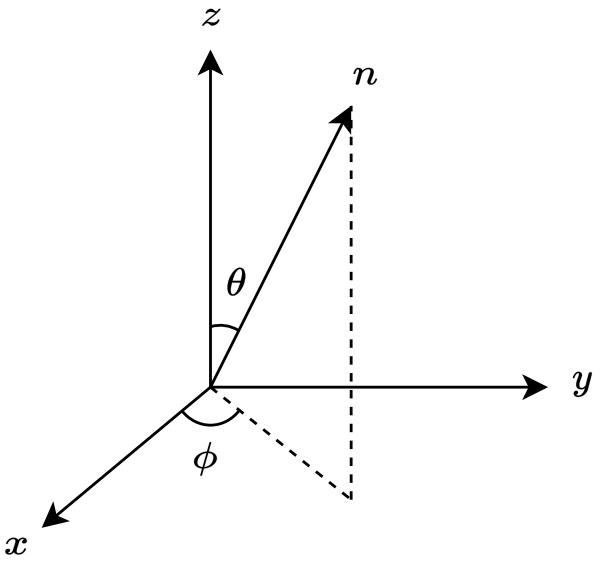


FIG. B.1. Schematic diagram of the definition of angles θ and ϕ.

The rotation matrix from the direction z to n is given by

$$\begin{aligned} R &= \begin{pmatrix} \sin\phi & \cos\phi & 0 \\ -\cos\phi & \sin\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix} \\ &= \begin{pmatrix} \sin\phi & \cos\theta\cos\phi & \sin\theta\cos\phi \\ -\cos\phi & \cos\theta\sin\phi & \sin\theta\sin\phi \\ 0 & -\sin\theta & \cos\theta \end{pmatrix}. \end{aligned} \tag{B.6}$$

Using this matrix, the quadrupole moments are transformed as

$$\ddot{Q}'_{ij} = \left(R^{\mathsf{T}}\ddot{Q}R\right)_{ij}, \tag{B.7}$$

and the plus and cross polarizations are then expressed as

$$h_+ = \frac{G}{c^4 r}(\ddot{Q}'_{xx} - \ddot{Q}'_{yy}), \tag{B.8}$$

$$h_\times = \frac{2G}{c^4 r}\ddot{Q}'_{xy}. \tag{B.9}$$

In practice, $\ddot{Q}'_{xx}$, $\ddot{Q}'_{yy}$, and $\ddot{Q}'_{xy}$ can be obtained through straightforward matrix calculations:

$$\ddot{Q}'_{xx} = \ddot{Q}_{xx}\sin^2\phi + \ddot{Q}_{yy}\cos^2\phi - \ddot{Q}_{xy}\sin 2\phi, \tag{B.10}$$

$$\begin{aligned} \ddot{Q}'_{yy} = {} & (\ddot{Q}_{xx}\cos^2\phi + \ddot{Q}_{yy}\sin^2\phi + \ddot{Q}_{xy}\sin 2\phi)\cos^2\theta + \ddot{Q}_{zz}\sin^2\theta \\ & - (\ddot{Q}_{xz}\cos\phi + \ddot{Q}_{yz}\sin\phi)\sin 2\theta, \end{aligned} \tag{B.11}$$

$$\ddot{Q}'_{xy} = \frac{1}{2}(\ddot{Q}_{xx} - \ddot{Q}_{yy})\cos\theta\sin 2\phi - \ddot{Q}_{xy}\cos\theta\cos 2\phi - (\ddot{Q}_{xz}\sin\phi - \ddot{Q}_{yz}\cos\phi)\sin\theta. \tag{B.12}$$

By substituting these expressions into Eqs. (B.8) and (B.9), we obtain the expressions

$$h_+ = \frac{G}{c^4 r}\{\ddot{Q}_{xx}\sin^2\phi + \ddot{Q}_{yy}\cos^2\phi - \ddot{Q}_{xy}\sin 2\phi + (\ddot{Q}_{xz}\cos\phi + \ddot{Q}_{yz}\sin\phi)\sin 2\theta \\ - \ddot{Q}_{zz}\sin^2\theta - (\ddot{Q}_{xx}\cos^2\phi + \ddot{Q}_{yy}\sin^2\phi + \ddot{Q}_{xy}\sin 2\phi)\cos^2\theta\}, \tag{B.13}$$

$$h_\times = \frac{G}{c^4 r}\{(\ddot{Q}_{xx} - \ddot{Q}_{yy})\cos\theta\sin 2\phi - 2\ddot{Q}_{xy}\cos\theta\cos 2\phi \\ - 2(\ddot{Q}_{xz}\sin\phi - \ddot{Q}_{yz}\cos\phi)\sin\theta\}. \tag{B.14}$$

Appendix C

Newtonian Approximation of Gravitational Waves from Binary Systems

Let us compute the inspiral GW signal of a compact binary system in a circular orbit. For simpility, we model each star as a point mass and use the Newtonian approximation.

The positions of the two point masses rotating in the xy-plane with an angular frequency ω_s and a distance R are given by

$$(x_1, y_1, z_1) = \left(-\frac{m_2}{m_1+m_2}R\sin(\omega_s t),\ \frac{m_2}{m_1+m_2}R\cos(\omega_s t),\ 0\right), \tag{C.1}$$

$$(x_2, y_2, z_2) = \left(\frac{m_1}{m_1+m_2}R\sin(\omega_s t),\ -\frac{m_1}{m_1+m_2}R\cos(\omega_s t),\ 0\right). \tag{C.2}$$

The components of the mass quadrupole moments for this system are then calculated as

$$Q_{xx} = m_1 x_1^2 + m_2 x_2^2 = \mu R^2 \frac{1-\cos(2\omega_s t)}{2}, \tag{C.3}$$

$$Q_{yy} = m_1 y_1^2 + m_2 y_2^2 = \mu R^2 \frac{1+\cos(2\omega_s t)}{2}, \tag{C.4}$$

$$Q_{xy} = m_1 x_1 y_1 + m_2 x_2 y_2 = -\mu R^2 \frac{\sin(2\omega_s t)}{2}, \tag{C.5}$$

$$Q_{xz} = Q_{yz} = Q_{zz} = 0. \tag{C.6}$$

Here we introduced the reduced mass μ, defined by

$$\mu = \frac{m_1 m_2}{m_1+m_2}. \tag{C.7}$$

Assuming that ω_s and R are constants, the second time derivatives of the quadrupole moments become

$$\ddot{Q}_{xx} = 2\mu R^2 \omega_s^2 \cos(2\omega_s t), \tag{C.8}$$

$$\ddot{Q}_{yy} = -2\mu R^2 \omega_s^2 \cos(2\omega_s t) = -\ddot{Q}_{xx}, \tag{C.9}$$

$$\ddot{Q}_{xy} = 2\mu R^2 \omega_s^2 \sin(2\omega_s t). \tag{C.10}$$

Substituting these results into Eqs. (B.13) and (B.14) yields

$$h_+(t) = \frac{1}{r}\frac{4G\mu\omega_s^2R^2}{c^4}\frac{1+\cos^2\theta}{2}\cos(2\omega_s(t-r/c)-2\phi) \tag{C.11}$$

$$h_\times(t) = -\frac{1}{r}\frac{4G\mu\omega_s^2R^2}{c^4}\cos\theta\sin(2\omega_s(t-r/c)-2\phi) \tag{C.12}$$

The angle θ is equivalent to the inclination angle (ι) between the orbit's normal direction and the line of sight. By shifting the time origin, certain terms in the equations can be simplified, resulting in

$$h_+(t) = \frac{1}{r}\frac{4G\mu\omega_s^2R^2}{c^4}\frac{1+\cos^2\iota}{2}\cos(2\omega_s t), \tag{C.13}$$

$$h_\times(t) = -\frac{1}{r}\frac{4G\mu\omega_s^2R^2}{c^4}\cos\iota\sin(2\omega_s t). \tag{C.14}$$

These equations indicate that the GW frequency is twice that of the binary system's rotation. Using Kepler's third law,

$$\omega_s^2 = \frac{G(m_1+m_2)}{R^3}, \tag{C.15}$$

and defining the chirp mass M_c,

$$M_c = \mu^{3/5}(m_1+m_2)^{2/5} = \frac{(m_1m_2)^{3/5}}{(m_1+m_2)^{1/5}}, \tag{C.16}$$

each mode is represented as

$$h_+(t) = \frac{4}{r}\left(\frac{GM_c}{c^2}\right)^{5/3}\left(\frac{\pi f_{\rm GW}}{c}\right)^{2/3}\frac{1+\cos^2\iota}{2}\cos(2\pi f_{\rm GW}t), \tag{C.17}$$

$$h_\times(t) = \frac{4}{r}\left(\frac{GM_c}{c^2}\right)^{5/3}\left(\frac{\pi f_{\rm GW}}{c}\right)^{2/3}\cos\iota\sin(2\pi f_{\rm GW}t), \tag{C.18}$$

where $2\pi f_{\rm GW} = 2\omega_s$. These expressions tell us that the GW amplitude depends on the chirp mass and not on the individual components in this approximation.

Up to this point, we have considered ω_s and R as constant values. Next, we will examine the influence of the changes in ω_s and R due to the energy radiation. The energy flux is calculated using Eq. (2.69):

$$\begin{aligned}\frac{{\rm d}E_{\rm GW}}{{\rm d}t} &= \frac{4G\mu^2\omega_s^6R^4}{\pi c^5}\int_0^\pi {\rm d}\theta\left\{\left(\frac{1+\cos^2\theta}{2}\right)^2+\cos^2\theta\right\}\sin\theta \\ &= \frac{32}{5}\frac{c^5}{G}\left(\frac{GM_c\omega_{\rm GW}}{2c^3}\right)^{10/3}.\end{aligned} \tag{C.19}$$

Differentiating Kepler's third law in Eq. (C.15) with respect to time, we obtain the relationship,

$$\dot{R} = -\frac{2}{3}R\omega_s\frac{\dot{\omega}_s}{\omega_s^2}. \tag{C.20}$$

Under the condition $\dot{\omega}_s \ll \omega_s^2$, $|\dot{R}|$ is significantly smaller than the orbital angular velocity

$\omega_s R$. This allows us to apply the approximation of a slowly changing circular orbit. The orbital energy is the sum of the kinetic and potential energies of the point masses, expressed as

$$E_{\rm orbit} = -\frac{Gm_1m_2}{2R} \tag{C.21}$$

$$= -\left(\frac{G^2M_c^5\omega_{\rm GW}^2}{32}\right)^{1/3} \tag{C.22}$$

Since the energy lost by the orbit is radiated as the GW, the energy flux of the GW can be expressed as

$$\frac{{\rm d}E_{\rm GW}}{{\rm d}t} = -\frac{{\rm d}E_{\rm orbit}}{{\rm d}t} = \frac{2}{3}\left(\frac{G^2M_c^5}{32}\right)^{1/3}\dot{\omega}_{\rm GW}\,\omega_{\rm GW}^{-1/3}. \tag{C.23}$$

Connecting this expression with the energy flux calculated from the quadrupole moments, as given in Eq. (C.19), we obtain the differential equation for $\omega_{\rm GW}$,

$$\dot{\omega}_{\rm GW} = \frac{12}{5}2^{1/3}\left(\frac{GM_c}{c^3}\right)^{5/3}\omega_{\rm GW}^{11/3}. \tag{C.24}$$

Solving this differential equation with the initial condition $\lim_{t\to t_{\rm coal}}\omega_{\rm GW}(t) = \infty$, we obtain the solution:

$$\omega_{\rm GW}(\tau) = \frac{1}{4}\left(\frac{5}{\tau}\right)^{3/8}\left(\frac{GM_c}{c^3}\right)^{-5/8} \tag{C.25}$$

Here, $t_{\rm coal}$ is the coalescence time, and $\tau = t_{\rm coal} - t$.

To derive the amplitude of the GW while taking account for energy radiation, we need to replace $\omega_s t$ in the particle positions in Eqs. (C.1) and (C.2) with

$$\Phi(t) = \int_{t_0}^{t} {\rm d}t'\, 2\omega_s(t')$$

$$= \int_{t_0}^{t} {\rm d}t'\, \omega_{\rm GW}(t'). \tag{C.26}$$

Substituting the obtained $\omega_{\rm GW}(\tau)$ and integrating, with $\tau = 0$ corresponding to $\Phi = \Phi_0$, we get

$$\Phi(\tau) = -2\left(\frac{5GM_c}{c^3}\right)^{-5/8}\tau^{5/8} + \Phi_0. \tag{C.27}$$

Under the condition $\dot{\omega}_s \ll \omega_s^2$, we can neglect the time derivatives of R and ω_s when computing the time derivatives of the quadrupole moments. Therefore, we simply need to replace the phase with Φ in Eqs. (C.17) and (C.18). Finally, we obtain

$$h_+(\tau) = \frac{1}{r}\left(\frac{GM_c}{c^2}\right)^{5/4}\left(\frac{5}{c\tau}\right)^{1/4}\left(\frac{1+\cos^2\iota}{2}\right)\cos\Phi(\tau), \tag{C.28}$$

$$h_\times(\tau) = -\frac{1}{r}\left(\frac{GM_c}{c^2}\right)^{5/4}\left(\frac{5}{c\tau}\right)^{1/4}\cos\iota\sin\Phi(\tau). \tag{C.29}$$

Figure C.1 plots the calculated plus and cross modes of the inspiral GW signal with $r = 400$ Mpc, $M_c = 30M_\odot$, $\iota = 0$, and $\Phi_0 = 0$. As the time approaches to the merger,

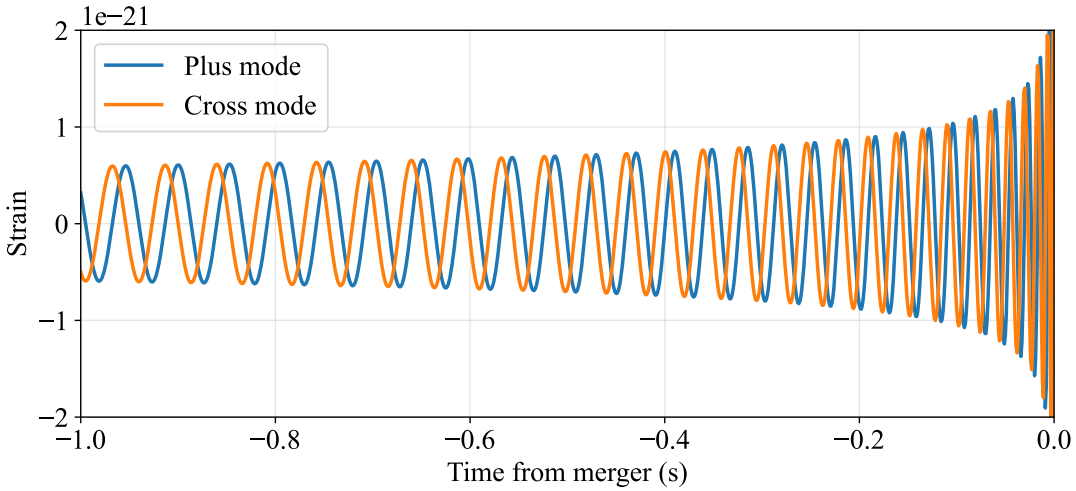


FIG. C.1. Amplitudes of the plus and cross modes of the inspiral signal derived from Newtonian approximation. We set $r = 400$ Mpc, $M_c = 30M_\odot$, $\iota = 0$, and $\Phi_0 = 0$.

both amplitude and frequency progressively increase. It is important to note that as the two stars in the binary system approach each other, the Newtonian approximation become less valid, and the effects of general relativity should be taken into account for accurately modeling GW signals.

Appendix D

Time-Frequency Uncertainty

Here, we derive the uncertainty relationship between time and frequency, following Ref. [255]. Consider a signal $x(t)$ with a finite energy E, given by

$$E = \int_{-\infty}^{\infty} \mathrm{d}t \ |x(t)|^2 < \infty. \tag{D.1}$$

The central time and frequency of $x(t)$ are defined as

$$\langle t \rangle = \frac{1}{E} \int_{-\infty}^{\infty} \mathrm{d}t \ t|x(t)|^2, \tag{D.2}$$

$$\langle \omega \rangle = \frac{1}{2\pi E} \int_{-\infty}^{\infty} \mathrm{d}\omega \ \omega|\tilde{x}(\omega)|^2, \tag{D.3}$$

respectively. Similarly, the duration σ_t and bandwidth σ_f of the signal are defined by

$$\sigma_t^2 = \frac{1}{E} \int_{-\infty}^{\infty} \mathrm{d}t \ (t - \langle t \rangle)^2|x(t)|^2, \tag{D.4}$$

$$\sigma_\omega^2 = \frac{1}{2\pi E} \int_{-\infty}^{\infty} \mathrm{d}\omega \ (\omega - \langle \omega \rangle)^2|\tilde{x}(\omega)|^2, \tag{D.5}$$

respectively. The uncertainty relationship between time and frequency is expressed as

$$\sigma_t \sigma_\omega \geq \frac{1}{2}. \tag{D.6}$$

To prove this relationship, we assume that the signal $x(t)$ is real, i.e. $x^*(t) = x(t)$, and the condition

$$\lim_{t \to \pm\infty} t|x(t)|^2 = 0 \tag{D.7}$$

holds. We can also assume $\langle t \rangle = \langle \omega \rangle = 0$ without loss of generality. Firstly, using the equation,

$$\dot{x}(t) = \frac{-i}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega \ \omega \tilde{x}(\omega) e^{-i\omega t}, \tag{D.8}$$

we have

$$\begin{aligned}\int_{-\infty}^{\infty} \mathrm{d}t\ |\dot{x}(t)|^2 &= \frac{1}{(2\pi)^2}\int_{-\infty}^{\infty} \mathrm{d}t\ e^{-i(\omega-\omega')t}\int_{-\infty}^{\infty} \mathrm{d}\omega\ \omega\tilde{x}(\omega)\int_{-\infty}^{\infty} \mathrm{d}\omega'\ \omega'\tilde{x}^*(\omega') \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\ \omega\tilde{x}(\omega)\int_{-\infty}^{\infty} \mathrm{d}\omega'\ \omega'\tilde{x}^*(\omega')\delta(\omega-\omega') \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\ \omega^2|\tilde{x}(\omega)|^2. \end{aligned} \tag{D.9}$$

Thus, we can evaluate $\sigma_t^2\sigma_\omega^2$ as

$$\sigma_t^2\sigma_\omega^2 = \frac{1}{E^2}\int_{-\infty}^{\infty} \mathrm{d}t\ |tx(t)|^2\int_{-\infty}^{\infty} \mathrm{d}t\ |\dot{x}(t)|^2 \geq \frac{1}{E^2}\left|\int_{-\infty}^{\infty} \mathrm{d}t\ tx(t)\dot{x}(t)\right|^2. \tag{D.10}$$

Here, we used the Cauchy-Schwarz inequality. The integrand of the right-hand side can be computed as

$$\begin{aligned}\int_{-\infty}^{\infty} \mathrm{d}t\ tx(t)\dot{x}(t) &= \frac{1}{2}\int_{-\infty}^{\infty} \mathrm{d}t\ t\frac{\mathrm{d}}{\mathrm{d}t}|x(t)|^2 \\ &= \frac{1}{2}\left\{\left[t|x(t)|^2\right]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \mathrm{d}t\ |x(t)|^2\right\} \\ &= -\frac{E}{2}. \end{aligned} \tag{D.11}$$

Therefore, we obtain

$$\sigma_t^2\sigma_\omega^2 \geq \frac{1}{4}. \tag{D.12}$$

Appendix E

Cubic Spline Interpolation

Cubic spline interpolation is a widely used method in numerical analysis to approximate a smooth curve passing through a set of given data points. Consider a data set consisting of $N+1$ data points: $(x_0, y_0), \ldots, (x_N, y_N)$. The objective of cubic spline interpolation is to construct a piecewise-defined function $f(x)$ that consists of N cubic polynomials, $f_i(x)$, defined over intervals $[x_{i-1}, x_i]$, where $i = 1, \ldots, N$. Each cubic polynomial $f_i(x)$ takes the form,

$$f_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i. \tag{E.1}$$

The coefficients are chosen to ensure that the interpolating function passes through all the given data points while maintaining continuity and smoothness. The four conditions that each cubic polynomial must satisfy are as follows:

$$f_i(x_{i-1}) = y_{i-1} \qquad \text{for } i = 1, \ldots, N \tag{E.2}$$

$$f_i(x_i) = y_i \qquad \text{for } i = 1, \ldots, N \tag{E.3}$$

$$f_i'(x_i) = f_{i+1}'(x_i) \qquad \text{for } i = 1, \ldots, N-1 \tag{E.4}$$

$$f_i''(x_i) = f_{i+1}''(x_i) \qquad \text{for } i = 1, \ldots, N-1 \tag{E.5}$$

These conditions specified for each cubic polynomial ensure the necessary continuity, smoothness, and interpolation through the given data points. However, to fully determine the $4N$ coefficients in total, two more equations are required. These additional equations typically involve boundary conditions to fix the behavior of the spline at the endpoints of the data set. A common type of boundary conditions used to complete the specification of the cubic spline is called the *natural boundary condition*, wherein the second derivatives at the endpoints of the data set are zero,

$$f_1''(x_0) = 0, \quad f_N''(x_N) = 0. \tag{E.6}$$

Now we have $4N$ equations that are necessary to determine all the coefficients. From Eqs. (E.2), (E.3), and (E.5), we obtain

$$a_i = \frac{v_i - v_{i-1}}{6h_i}, \tag{E.7}$$

$$b_i = \frac{v_i}{2}, \tag{E.8}$$

$$c_i = \frac{h_i}{6}(2v_i + v_{i-1}) + \frac{y_i - y_{i-1}}{h_i}, \tag{E.9}$$

$$d_i = y_i, \tag{E.10}$$

where $v_i = f_i''(x_i)$ and $h_i = x_i - x_{i-1}$ for $i = 1, \ldots, N$, and $v_0 = f_1''(x_0)$. Inserting these into Eq. (E.4) results in

$$h_i v_{i-1} + 2(h_i + h_{i+1})v_i + h_{i+1}v_{i+1} = w_i, \tag{E.11}$$

where

$$w_i = 6\left(\frac{y_{i+1} - y_i}{h_{i+1}} - \frac{y_i - y_{i-1}}{h_i}\right) \tag{E.12}$$

for $i = 1, \ldots, N-1$. From the natural boundary conditions Eq. (E.6), we have $v_0 = v_N = 0$. The equations for $\{v_i\}$ can be represented in matrix form as

$$\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ h_1 & 2(h_1+h_2) & h_2 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & h_{N-1} & 2(h_{N-1}+h_N) & h_N \\ 0 & \cdots & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_0 \\ v_1 \\ \vdots \\ v_{N-1} \\ v_N \end{pmatrix} = \begin{pmatrix} 0 \\ w_1 \\ \vdots \\ w_{N-1} \\ 0 \end{pmatrix} \tag{E.13}$$

Computing the coefficients $\{a_i, b_i, c_i, d_i\}$ is now equivalent to solving this matrix equation for $v_0, \ldots, v_N$. The coefficient matrix on the left side of this equation is a tridiagonal matrix, and this can be easily solved by computers. Once this matrix equation is solved, we can obtain the interpolating functions by substituting the values of $v_0, \ldots, v_N$ into Eqs. (E.7)-(E.10).

Appendix F

Solution of Cubic Equations

Cubic equations have analytical solutions, known as Cardano's formulae. The analytical solutions of cubic equations in the form $a_3x^3 + a_2x^2 + a_1x + a_0 = 0$ are expressed as

$$x_k = -\frac{a_2}{3a_3} + \omega^k \sqrt[3]{-q + \sqrt{q^2 + p^3}} + \omega^{3-k} \sqrt[3]{-q - \sqrt{q^2 + p^3}} \quad (k = 0, 1, 2), \tag{F.1}$$

where ω is a complex cube root of unity, and p and q are defined by

$$p = \frac{3a_3a_1 - a_2^2}{9a_3^2}, \tag{F.2}$$

$$q = \frac{27a_3^2a_0 - 9a_3a_2a_1 + 2a_2^3}{54a_3^3}. \tag{F.3}$$

While Cardano's formulae provide exact solutions, numerical methods are often preferred. One can construct a *companion matrix* C associated with the cubic equation $a_3x^3 + a_2x^2 + a_1x + a_0 = 0$ as

$$C = \begin{pmatrix} 0 & 0 & -a_0/a_3 \\ 1 & 0 & -a_1/a_3 \\ 0 & 1 & -a_2/a_3 \end{pmatrix}. \tag{F.4}$$

The eigenvalues λ of this matrix satisfy

$$\det(\lambda I - C) = \lambda^3 + \frac{a_2}{a_3}\lambda^2 + \frac{a_1}{a_3}\lambda + \frac{a_0}{a_3} = 0. \tag{F.5}$$

Therefore, the eigenvalues of the companion matrix are the solutions of the cubic equation. Efficient numerical methods, such as those provided by LAPACK [256], can be employed to find these eigenvalues.

References

[1] S. Sasaoka, Y. Hou, K. Somiya, and H. Takahashi, Localization of gravitational waves using machine learning, Phys. Rev. D **105**, 103030 (2022).

[2] S. Sasaoka, N. Koyama, D. Dominguez, Y. Sakai, K. Somiya, Y. Omae, and H. Takahashi, Visualizing convolutional neural network for classifying gravitational waves from core-collapse supernovae, Phys. Rev. D **108**, 123033 (2023).

[3] S. Sasaoka, N. Koyama, D. Dominguez, Y. Sakai, K. Somiya, Y. Omae, and H. Takahashi, Comparative study of 1D and 2D convolutional neural network models with attribution analysis for gravitational wave detection from compact binary coalescences, Phys. Rev. D **109**, 043011 (2024).

[4] J. P. Gardner *et al.*, The James Webb Space Telescope, Space Sci. Rev. **123**, 485 (2006).

[5] J. Aasi *et al.* (LIGO Scientific Collaboration), Advanced LIGO, Classical Quantum Gravity **32**, 074001 (2015).

[6] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Observation of gravitational waves from a binary black hole merger, Phys. Rev. Lett. **116**, 061102 (2016).

[7] A. Einstein, Näherungsweise integration der feldgleichungen der gravitation, Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) , 688 (1916).

[8] A. Einstein, Über gravitationswellen, Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.) , 154 (1918).

[9] J. Weber, Detection and generation of gravitational waves, Phys. Rev. **117**, 306 (1960).

[10] J. Weber, Evidence for discovery of gravitational radiation, Phys. Rev. Lett. **22**, 1320 (1969).

[11] J. A. Tyson, Null search for bursts of gravitational radiation, Phys. Rev. Lett. **31**, 326 (1973).

[12] R. A. Hulse and J. H. Taylor, Discovery of a pulsar in a binary system., Astrophys. J. **195**, L51 (1975).

[13] J. H. Taylor, L. A. Fowler, and P. M. McCulloch, Measurements of general relativistic effects in the binary pulsar PSR1913 + 16, Nature **277**, 437 (1979).

[14] M. Ando *et al.* (TAMA Collaboration), Stable operation of a 300-m laser interferometer with sufficient sensitivity to detect gravitational-wave events within our galaxy, Phys. Rev. Lett. **86**, 3950 (2001).

[15] B. Willke *et al.*, The GEO 600 gravitational wave detector, Classical Quantum Gravity **19**, 1377 (2002).

[16] T. Accadia *et al.*, Virgo: A laser interferometer to detect gravitational waves, J. Instrum. **7**, 3012 (2012).

[17] A. Abramovici *et al.*, LIGO: The laser interferometer gravitational-wave observatory, Science **256**, 325 (1992).

[18] F. Acernese *et al.*, Advanced Virgo: A second-generation interferometric gravitational wave detector, Classical Quantum Gravity **32**, 024001 (2014).

[19] T. Akutsu *et al.*, KAGRA: 2.5 generation interferometric gravitational wave detector, Nat. Astron. **3**, 35 (2019).

[20] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), GW170817: Observation of gravitational waves from a binary neutron star inspiral, Phys. Rev. Lett. **119**, 161101 (2017).

[21] B. P. Abbott *et al.*, Multi-messenger observations of a binary neutron star merger, Astrophys. J. Lett. **848**, L12 (2017).

[22] A. Goldstein *et al.*, An ordinary short gamma-ray burst with extraordinary implications: Fermi-GBM detection of GRB 170817A, Astrophys. J. Lett. **848**, L14 (2017).

[23] V. Savchenko *et al.*, INTEGRAL detection of the first prompt gamma-ray signal coincident with the gravitational-wave event GW170817, Astrophys. J. Lett. **848**, L15 (2017).

[24] D. Radice, A. Perego, F. Zappa, and S. Bernuzzi, GW170817: Joint constraint on the neutron star equation of state from multimessenger observations, Astrophys. J. Lett. **852**, L29 (2018).

[25] B. P. Abbott *et al.* (The LIGO Scientific Collaboration and the Virgo Collaboration), GW170817: Measurements of neutron star radii and equation of state, Phys. Rev. Lett. **121**, 161101 (2018).

[26] T. Baker, E. Bellini, P. G. Ferreira, M. Lagos, J. Noller, and I. Sawicki, Strong constraints on cosmological gravity from GW170817 and GRB 170817A, Phys. Rev. Lett. **119**, 251301 (2017).

[27] J. M. Ezquiaga and M. Zumalacárregui, Dark energy after GW170817: Dead ends and the road ahead, Phys. Rev. Lett. **119**, 251304 (2017).

[28] B. P. Abbott *et al.*, A gravitational-wave standard siren measurement of the Hubble constant, Nature **551**, 85 (2017).

[29] S. J. Smartt *et al.*, A kilonova as the electromagnetic counterpart to a gravitational-wave source, Nature **551**, 75 (2017).

[30] R. Abbott *et al.*, Observation of gravitational waves from two neutron star–black hole coalescences, Astrophys. J. Lett. **915**, L5 (2021).

[31] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), 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**, 031040 (2019).

[32] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), 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).

[33] R. Abbott *et al.* (The LIGO Scientific Collaboration and the Virgo Collaboration), GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run, Phys. Rev. D **109**, 022001 (2024).

[34] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run, Phys. Rev. X **13**, 041039 (2023).

[35] https://gracedb.ligo.org/superevents/public/O4.

[36] LIGO Scientific Collaboration, Data release for event GW150914.

[37] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, FINDCHIRP: An algorithm for detection of gravitational waves from inspiraling compact binaries, Phys. Rev. D **85**, 122006 (2012).

[38] R. M. Bionta *et al.*, Observation of a neutrino burst in coincidence with supernova 1987A in the large magellanic cloud, Phys. Rev. Lett. **58**, 1494 (1987).

[39] K. Hirata *et al.*, Observation of a neutrino burst from the supernova SN1987A, Phys. Rev. Lett. **58**, 1490 (1987).

[40] K. Kotake, W. Iwakami, N. Ohnishi, and S. Yamada, Stochastic nature of gravitational waves from supernova explosions with standing accretion shock instability, Astrophys. J. **697**, L133 (2009).

[41] E. Cuoco *et al.*, Enhancing gravitational-wave science with machine learning, Mach. Learn. **2**, 011002 (2020).

[42] T. Zhao, R. Shi, Y. Zhou, Z. Cao, and Z. Ren, Dawning of a new era in gravitational wave data analysis: Unveiling cosmic mysteries via artificial intelligence – A systematic review, arXiv:2311.15585.

[43] L. Alzubaidi *et al.*, Review of deep learning: concepts, CNN architectures, challenges, applications, future directions, J. Big Data **8**, 53 (2021).

[44] D. George and E. A. Huerta, Deep neural networks to enable real-time multimessenger astrophysics, Phys. Rev. D **97**, 044039 (2018).

[45] P. G. Krastev, Real-time detection of gravitational waves from binary neutron stars using artificial neural networks, Phys. Lett. B **803**, 135330 (2020).

[46] 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**, 122002 (2018).

[47] M. López, I. Di Palma, M. Drago, P. Cerdá-Durán, and F. Ricci, Deep learning for core-collapse supernova detection, Phys. Rev. D **103**, 063011 (2021).

[48] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith, Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy, Nat. Phys. **18**, 112 (2021).

[49] M. Dax, S. R. Green, J. Gair, J. H. Macke, A. Buonanno, and B. Schölkopf, Real-time gravitational wave science with neural posterior estimation, Phys. Rev. Lett. **127**, 241103 (2021).

[50] M. Dax, S. R. Green, J. Gair, M. Pürrer, J. Wildberger, J. H. Macke, A. Buonanno, and B. Schölkopf, Neural importance sampling for rapid and reliable gravitational-wave inference, Phys. Rev. Lett. **130**, 171403 (2023).

[51] M. Zevin *et al.*, Gravity Spy: Integrating Advanced LIGO detector characterization, machine learning, and citizen science, Classical Quantum Gravity **34**, 064003 (2017).

[52] S. Bahaadini, N. Rohani, S. Coughlin, M. Zevin, V. Kalogera, and A. K. Katsaggelos, Deep multi-view models for glitch classification, in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, New Orleans, LA, USA, 2017) pp. 2931–2935.

[53] S. Bahaadini, V. Noroozi, N. Rohani, S. Coughlin, M. Zevin, J. Smith, V. Kalogera, and A. Katsaggelos, Machine learning for Gravity Spy: Glitch classification and dataset, Inf. Sci. **444**, 172 (2018).

[54] N. Mukund *et al.*, Neural sensing and control in a kilometer-scale gravitational-wave observatory, Phys. Rev. Appl. **20**, 064041 (2023).

[55] A. B. Arrieta *et al.*, Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion **58**, 82 (2020).

[56] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. A **454**, 903 (1998).

[57] M. Takeda *et al.*, Application of the Hilbert-Huang transform for analyzing standing-accretion-shock-instability induced gravitational waves in a core-collapse supernova, Phys. Rev. D **104**, 084063 (2021).

[58] M. Maggiore, *Gravitational Waves: Volume 1: Theory and Experiments* (Oxford University Press, Oxford, 2007).

[59] *Latest estimated sensitivity of KAGRA (v201708)*, Tech. Rep. T1707038-v9 (JGW Document, 2017).

[60] D. Shoemaker, R. Schilling, L. Schnupp, W. Winkler, K. Maischberger, and A. Rüdiger, Noise behavior of the Garching 30-meter prototype gravitational-wave detector, Phys. Rev. D **38**, 423 (1988).

[61] K. Somiya, Detector configuration of KAGRA–the Japanese cryogenic gravitational-wave detector, Classical Quantum Gravity **29**, 124007 (2012).

[62] V. B. Braginsky, Classical and quantum restrictions on the detection of weak disturbances of a macroscopic oscillator, Sov. Phys. JETP **26**, 831 (1968).

[63] J. Aasi *et al.*, Enhanced sensitivity of the LIGO gravitational wave detector by using squeezed states of light, Nat. Photon. **7**, 613 (2013).

[64] C. S. Unnikrishnan, IndIGO and LIGO-India: Scope and plans for gravitational wave research and precision metrology in India, Int. J. Mod. Phys. D **22**, 1341010 (2013).

[65] K. Ackley *et al.*, Neutron star extreme matter observatory: A kilohertz-band gravitational-wave detector in the global network, Pub. Astron. Soc. Aust. **37**, e047 (2020).

[66] M. Punturo *et al.*, The Einstein Telescope: A third-generation gravitational wave observatory, Classical Quantum Gravity **27**, 194002 (2010).

[67] M. Branchesi *et al.*, Science with the Einstein Telescope: A comparison of different designs, J. Cosmol. Astropart. Phys. **2023**, 068 (2023).

[68] B. P. Abbott *et al.*, Exploring the sensitivity of next generation gravitational wave detectors, Classical Quantum Gravity **34**, 044001 (2017).

[69] D. Reitze *et al.*, Cosmic Explorer: The U.S. contribution to gravitational-wave astronomy beyond LIGO, Bull. Am. Astron. Soc. **51**, 035 (2019).

[70] M. Maggiore *et al.*, Science case for the Einstein Telescope, J. Cosmol. Astropart. Phys. **2020**, 050 (2020).

[71] https://www.gwoptics.org/research/et/layout.

[72] P. Amaro-Seoane *et al.*, Laser interferometer space antenna, arXiv:1702.00786.

[73] S. Kawamura *et al.*, The Japanese space gravitational wave antenna: DECIGO, Classical Quantum Gravity **28**, 094011 (2011).

[74] J. Luo *et al.*, TianQin: A space-borne gravitational wave detector, Classical Quantum Gravity **33**, 035010 (2016).

[75] W.-R. Hu and Y.-L. Wu, The Taiji program in space for gravitational wave physics and the nature of gravity, Natl. Sci. Rev. **4**, 685 (2017).

[76] C. J. Moore, R. H. Cole, and C. P. L. Berry, Gravitational-wave sensitivity curves, Classical Quantum Gravity **32**, 015014 (2014).

[77] http://gwplotter.com.

[78] R. C. Tolman, Static solutions of Einstein's field equations for spheres of fluid, Phys. Rev. **55**, 364 (1939).

[79] J. R. Oppenheimer and G. M. Volkoff, On massive neutron cores, Phys. Rev. **55**, 374 (1939).

[80] B. R. Iyer and C. M. Will, Post-Newtonian gravitational radiation reaction for two-body systems, Phys. Rev. Lett. **70**, 113 (1993).

[81] B. R. Iyer and C. M. Will, Post-Newtonian gravitational radiation reaction for two-body systems: Nonspinning bodies, Phys. Rev. D **52**, 6882 (1995).

[82] F. Pretorius, Evolution of binary black-hole spacetimes, Phys. Rev. Lett. **95**, 121101 (2005).

[83] M. Campanelli, C. O. Lousto, P. Marronetti, and Y. Zlochower, Accurate evolutions of orbiting black-hole binaries without excision, Phys. Rev. Lett. **96**, 111101 (2006).

[84] J. G. Baker, J. Centrella, D.-I. Choi, M. Koppitz, and J. van Meter, Gravitational-wave extraction from an inspiraling configuration of merging black holes, Phys. Rev. Lett. **96**, 111102 (2006).

[85] A. Buonanno and T. Damour, Effective one-body approach to general relativistic two-body dynamics, Phys. Rev. D **59**, 084006 (1999).

[86] A. 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**, 044028 (2017).

[87] P. Ajith *et al.*, A phenomenological template family for black-hole coalescence waveforms, Classical Quantum Gravity **24**, S689 (2007).

[88] G. Pratten *et al.*, Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes, Phys. Rev. D **103**, 104056 (2021).

[89] *LIGO-Virgo-KAGRA Cumulative Detection plot - O1-O4a*, Tech. Rep. G2302098-v11 (LIGO Document, 2024).

[90] K. Kotake, K. Sato, and K. Takahashi, Explosion mechanism, neutrino burst and gravitational wave in core-collapse supernovae, Rep. Prog. Phys. **69**, 971 (2006).

[91] H.-T. Janka, Explosion mechanisms of core-collapse supernovae, Annu. Rev. Nucl. Part. Sci. **62**, 407 (2012).

[92] E. Abdikamalov, G. Pagliaroli, and D. Radice, Gravitational waves from core-collapse supernovae, in *Handbook of Gravitational Wave Astronomy* (Springer Singapore, Singapore, 2020) pp. 1–37.

[93] L. F. Roberts, G. Shen, V. Cirigliano, J. A. Pons, S. Reddy, and S. E. Woosley, Protoneutron star cooling with convection: The effect of the symmetry energy, Phys. Rev. Lett. **108**, 061103 (2012).

[94] J. M. Blondin, A. Mezzacappa, and C. DeMarino, Stability of standing accretion shocks, with an eye toward core-collapse supernovae, Astrophys. J. **584**, 971 (2003).

[95] T. Kuroda, K. Kotake, and T. Takiwaki, A new gravitational-wave signature from standing accretion shock instability in supernovae, Astrophys. J. Lett. **829**, L14 (2016).

[96] P. D. Lasky, Gravitational waves from neutron stars: A review, Pub. Astron. Soc. Aust. **32** (2015), 10.1017/pasa.2015.35.

[97] K. Riles, Searches for continuous-wave gravitational radiation, Liv. Rev. Rel. **26** (2023).

[98] R. Abbott *et al.* (LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration), All-sky search for continuous gravitational waves from isolated neutron stars using Advanced LIGO and Advanced Virgo O3 data, Phys. Rev. D **106**, 102008 (2022).

[99] K. Wette, Searches for continuous gravitational waves from neutron stars: A twenty-year retrospective, Astrop. Phys. **153**, 102880 (2023).

[100] N. Christensen, Stochastic gravitational wave backgrounds, Rep. Prog. Phys. **82**, 016903 (2018).

[101] E. Thrane, N. Christensen, and R. M. S. Schofield, Correlated magnetic noise in global networks of gravitational-wave detectors: Observations and implications, Phys. Rev. D **87**, 123009 (2013).

[102] S. Atsuta *et al.*, Measurement of Schumann resonance at Kamioka, J. Phys. Conf. Ser. **716**, 012020 (2016).

[103] B. Allen and J. D. Romano, Detecting a stochastic background of gravitational radiation: Signal processing strategies and sensitivities, Phys. Rev. D **59**, 102001 (1999).

[104] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Upper limits on the stochastic gravitational-wave background from Advanced LIGO's first observing run, Phys. Rev. Lett. **118**, 121101 (2017).

[105] J. D. E. Creighton and W. G. Anderson, *Gravitational-Wave Physics and Astronomy: An Introduction to Theory, Experiment and Data Analysis* (Wiley-VCH, New York, 2011).

[106] J. Neyman and E. S. Pearson, On the problem of the most efficient tests of statistical hypotheses, Phil. Trans. R. Soc. A **231**, 289 (1933).

[107] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. **21**, 1087 (1953).

[108] J. Skilling, Nested sampling, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, Vol. 735 (AIP, Garching, Germany, 2004) pp. 395–405.

[109] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika **57**, 97 (1970).

[110] S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-6**, 721 (1984).

[111] D. Gabor, Theory of communication. Part 1: The analysis of information, J. Inst. Electr. Eng. **93**, 429 (1946).

[112] J. C. Brown, Calculation of a constant Q spectral transform, J. Acoust. Soc. Am. **89**, 425 (1991).

[113] S. Chatterji, L. Blackburn, G. Martin, and E. Katsavounidis, Multiresolution techniques for the detection of gravitational-wave bursts, Classical Quantum Gravity **21**, S1809 (2004).

[114] S. K. Chatterji, *The search for gravitational wave bursts in data from the second LIGO science run*, Ph.D. thesis, Massachusetts Institute of Technology (2005).

[115] V. Nair and G. E. Hinton, Rectified linear units improve restricted Boltzmann machines, in *Proceedings of the 27th International Conference on International Conference on Machine Learning* (Omnipress, Madison, WI, USA, 2010) pp. 807–814.

[116] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, J. Mach. Learn. Res. **15**, 1929 (2014).

[117] S. Ioffe and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in *Proceedings of the 32nd International Conference on Machine Learning*, Vol. 37 (PMLR, Lille, France, 2015) pp. 448–456.

[118] H. Robbins and S. Monro, A stochastic approximation method, Ann. Math. Statist. **22**, 400 (1951).

[119] S. Ruder, An overview of gradient descent optimization algorithms, arXiv:1609.04747.

[120] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, On the importance of initialization and momentum in deep learning, in *Proceedings of the 30th International Conference on Machine Learning* (PMLR, Atlanta, Georgia, USA, 2013) pp. 1139–1147.

[121] J. Duchi, E. Hazan, and Y. Singer, Adaptive subgradient methods for online learning and stochastic optimization, J. Mach. Learn. Res. **12**, 2121 (2011).

[122] G. Hinton, Neural networks for machine learning lecture notes.

[123] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv:1412.6980.

[124] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM **60**, 84 (2017).

[125] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, ImageNet: A large-scale hierarchical image database, in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, Miami, FL, USA, 2009) pp. 248–255.

[126] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556.

[127] K. He, X. Zhang, S. Ren, and J. Sun, Deep residual learning for image recognition, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Las Vegas, NV, USA, 2016) pp. 770–778.

[128] M. Sundararajan, A. Taly, and Q. Yan, Axiomatic attribution for deep networks, in *Proceedings of the 34th International Conference on Machine Learning* (PMLR, Sydney, NSW, Australia, 2017) pp. 3319–3328.

[129] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, Learning deep features for discriminative localization, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Las Vegas, NV, USA, 2016) pp. 2921–2929.

[130] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization, in *2017 IEEE International Conference on Computer Vision (ICCV)* (IEEE, Venice, Italy, 2017) pp. 618–626.

[131] A. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks, in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)* (IEEE, Lake Tahoe, NV, USA, 2018) pp. 839–847.

[132] H. Wang, Z. Wang, M. Du, F. Yang, Z. Zhang, S. Ding, P. Mardziel, and X. Hu, Score-CAM: Score-weighted visual explanations for convolutional neural networks, in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (IEEE, Seattle, WA, USA, 2020) pp. 111–119.

[133] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, Striving for simplicity: The all convolutional net, arXiv:1412.6806.

[134] H. Gabbard, M. Williams, F. Hayes, and C. Messenger, Matching matched filtering with deep networks for gravitational-wave astronomy, Phys. Rev. Lett. **120**, 141103 (2018).

[135] T. D. Gebhard, N. Kilbertus, I. Harry, and B. Schölkopf, Convolutional neural networks: A magic bullet for gravitational-wave detection?, Phys. Rev. D **100**, 063015 (2019).

[136] H. Wang, S. Wu, Z. Cao, X. Liu, and J.-Y. Zhu, Gravitational-wave signal recognition of LIGO data by deep learning, Phys. Rev. D **101**, 104003 (2020).

[137] J. Yan, M. Avagyan, R. E. Colgan, D. b. u. Veske, I. Bartos, J. Wright, Z. Márka, and S. Márka, Generalized approach to matched filtering using neural networks, Phys. Rev. D **105**, 043006 (2022).

[138] M. B. Schäfer *et al.*, First machine learning gravitational-wave search mock data challenge, Phys. Rev. D **107**, 023021 (2023).

[139] A. H. Nitz, T. Dent, T. D. Canton, S. Fairhurst, and D. A. Brown, Detecting binary compact-object mergers with gravitational waves: Understanding and improving the sensitivity of the PyCBC search, Astrophys. J. **849**, 118 (2017).

[140] P. Nousi, A. E. Koloniari, N. Passalis, P. Iosif, N. Stergioulas, and A. Tefas, Deep residual networks for gravitational wave detection, Phys. Rev. D **108**, 024022 (2023).

[141] P. G. Krastev, K. Gill, V. A. Villar, and E. Berger, Detection and parameter estimation of gravitational waves from binary neutron-star mergers in real LIGO data using deep learning, Phys. Lett. B **815**, 136161 (2021).

[142] W. Wei and E. Huerta, Deep learning for gravitational wave forecasting of neutron star mergers, Phys. Lett. B **816**, 136185 (2021).

[143] G. Baltus, J. Janquart, M. Lopez, A. Reza, S. Caudill, and J.-R. Cudell, Convolutional neural networks for the detection of the early inspiral of a gravitational-wave signal, Phys. Rev. D **103**, 102003 (2021).

[144] J. a. Aveiro, F. F. Freitas, M. Ferreira, A. Onofre, C. m. c. Providência, G. m. c. Gonçalves, and J. A. Font, Identification of binary neutron star mergers in gravitational-wave data using object-detection machine learning models, Phys. Rev. D **106**, 084059 (2022).

[145] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, You only look once: Unified, real-time object detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, Las Vegas, NV, USA, 2016) pp. 779–788.

[146] W. Wei, E. A. Huerta, M. Yun, N. Loutrel, M. A. Shaikh, P. Kumar, R. Haas, and V. Kindratenko, Deep learning with quantized neural networks for gravitational-wave forecasting of eccentric compact binary coalescence, Astrophys. J. **919**, 82 (2021).

[147] R. Qiu, P. G. Krastev, K. Gill, and E. Berger, Deep learning detection and classification of gravitational waves from neutron star-black hole mergers, Phys. Lett. B **840**, 137850 (2023).

[148] S. Klimenko and G. Mitselmakher, A wavelet method for detection of gravitational wave bursts, Classical Quantum Gravity **21**, S1819 (2004).

[149] S. Klimenko *et al.*, Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors, Phys. Rev. D **93**, 042004 (2016).

[150] A. Iess, E. Cuoco, F. Morawski, and J. Powell, Core-collapse supernova gravitational-wave search and deep learning classification, Mach. Learn. **1**, 025014 (2020).

[151] M. L. Chan, I. S. Heng, and C. Messenger, Detection and classification of supernova gravitational wave signals: A deep learning approach, Phys. Rev. D **102**, 043022 (2020).

[152] A. Iess, E. Cuoco, F. Morawski, C. Nicolaou, and O. Lahav, LSTM and CNN application for core-collapse supernova search in gravitational wave real data, Astron. Astrophys. **669**, A42 (2023).

[153] M. C. Edwards, Classifying the equation of state from rotating core collapse gravitational waves with deep learning, Phys. Rev. D **103**, 024025 (2021).

[154] Y.-S. Chao, C.-Z. Su, T.-Y. Chen, D.-W. Wang, and K.-C. Pan, Determining the core structure and nuclear equation of state of rotating core-collapse supernovae with gravitational waves by convolutional neural networks, Astrophys. J. **939**, 13 (2022).

[155] J. Powell, A. Iess, M. Llorens-Monteagudo, M. Obergaulinger, B. Müller, A. Torres-Forné, E. Cuoco, and J. A. Font, Determining the core-collapse supernova explosion mechanism with current and future gravitational-wave observatories, arXiv:2311.18221.

[156] J. Bayley, C. Messenger, and G. Woan, Rapid parameter estimation for an all-sky continuous gravitational wave search using conditional variational auto-encoders, Phys. Rev. D **106**, 083022 (2022).

[157] T. S. Yamamoto and T. Tanaka, Use of an excess power method and a convolutional neural network in an all-sky search for continuous gravitational waves, Phys. Rev. D **103**, 084049 (2021).

[158] C. Dreissigacker, R. Sharma, C. Messenger, R. Zhao, and R. Prix, Deep-learning continuous gravitational waves, Phys. Rev. D **100**, 044009 (2019).

[159] C. Dreissigacker and R. Prix, Deep-learning continuous gravitational waves: Multiple detectors and realistic noise, Phys. Rev. D **102**, 022005 (2020).

[160] P. M. Joshi and R. Prix, Novel neural-network architecture for continuous gravitational waves, Phys. Rev. D **108**, 063021 (2023).

[161] A. Utina, F. Marangio, F. Morawski, A. Iess, T. Regimbau, G. Fiameni, and E. Cuoco, Deep learning searches for gravitational wave stochastic backgrounds, in *2021 International Conference on Content-Based Multimedia Indexing (CBMI)* (IEEE, Lille, France, 2021) pp. 1–6.

[162] T. S. Yamamoto, S. Kuroyanagi, and G.-C. Liu, Deep learning for intermittent gravitational wave signals, Phys. Rev. D **107**, 044032 (2023).

[163] C. Chatterjee, L. Wen, K. Vinsen, M. Kovalam, and A. Datta, Using deep learning to localize gravitational wave sources, Phys. Rev. D **100**, 103025 (2019).

[164] D. J. Rezende and S. Mohamed, Variational inference with normalizing flows, in *Proceedings of the 32nd International Conference on Machine Learning* (PMLR, Lille, France, 2015) pp. 1530–1538.

[165] C. Chatterjee, M. Kovalam, L. Wen, D. Beveridge, F. Diakogiannis, and K. Vinsen, Rapid localization of gravitational wave sources from compact binary coalescences using deep learning, Astrophys. J. **959**, 42 (2023).

[166] C. Chatterjee and L. Wen, Premerger sky localization of gravitational waves from binary neutron star mergers using deep learning, Astrophys. J. **959**, 76 (2023).

[167] F. Robinet, N. Arnaud, N. Leroy, A. Lundgren, D. Macleod, and J. McIver, Omicron: A tool to characterize transient noise in gravitational-wave detectors, SoftwareX **12**, 100620 (2020).

[168] J. Powell, L. Sun, K. Gereb, P. D. Lasky, and M. Dollmann, Generating transient noise artefacts in gravitational-wave detector data with generative adversarial networks, Classical Quantum Gravity **40**, 035006 (2023).

[169] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, Generative adversarial nets, in *Advances in Neural Information Processing Systems* (Curran Associates, Montreal, Quebec, Canada, 2014) pp. 2672–2680.

[170] K. Sakai, K. Oohara, M. Kaneyama, and H. Takahashi, Analysis of the real gravitational wave data GW150914 with the Hilbert-Huang transform, ICIC Exp. Lett. **11**, 45 (2017).

[171] K. Sakai, K. Oohara, H. Nakano, M. Kaneyama, and H. Takahashi, Estimation of starting times of quasinormal modes in ringdown gravitational waves with the Hilbert-Huang transform, Phys. Rev. D **96**, 044047 (2017).

[172] L. Cohen, *Time-frequency analysis*, Vol. 778 (Prentice hall New Jersey, 1995).

[173] I. Yoda, K. Oohara, H. Takahashi, and K. Sakai, Precise analysis of gravitational waves from binary neutron star coalescence using Hilbert-Huang transform based on Akima spline interpolation, Prog. Theor. Exp. Phys. **2023**, 083E01 (2023).

[174] Z. Wu and N. E. Huang, Ensemble empirical mode decomposition: A noise-assisted data analysis method, Adv. Adapt. Data Anal. **01**, 1 (2009).

[175] J.-R. Yeh, J.-S. Shieh, and N. E. Huang, Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method, Adv. Adapt. Data Anal. **02**, 135 (2010).

[176] J. B. Camp, J. K. Cannizzo, and K. Numata, Application of the Hilbert-Huang transform to the search for gravitational waves, Phys. Rev. D **75**, 061101 (2007).

[177] A. Stroeer, J. K. Cannizzo, J. B. Camp, and N. Gagarin, Methods for detection and characterization of signals in noisy data with the Hilbert-Huang transform, Phys. Rev. D **79**, 124022 (2009).

[178] A. Stroeer, L. Blackburn, and J. Camp, Comparison of signals from gravitational wave detectors with instantaneous time-frequency maps, Classical Quantum Gravity **28**, 155001 (2011).

[179] H. Takahashi, K. Oohara, M. Kaneyama, Y. Hiranuma, and J. B. Camp, On investigating EMD parameters to search for gravitational waves, Adv. Adapt. Data Anal. **05**, 1350010 (2013).

[180] A. Akhshi, H. Alimohammadi, S. Baghram, S. Rahvar, M. R. R. Tabar, and H. Arfaei, A template-free approach for waveform extraction of gravitational wave events, Sci. Rep. **11** (2021).

[181] C.-P. Hu, L. C.-C. Lin, K.-C. Pan, K.-L. Li, C.-C. Yen, A. K. H. Kong, and C. Y. Hui, A comprehensive analysis of the gravitational wave events with the stacked Hilbert-Huang transform: From compact binary coalescence to supernova, Astrophys. J. **935**, 127 (2022).

[182] M. Kaneyama, K. Oohara, H. Takahashi, Y. Sekiguchi, H. Tagoshi, and M. Shibata, Analysis of gravitational waves from binary neutron star merger by Hilbert-Huang transform, Phys. Rev. D **93**, 123010 (2016).

[183] H. Akima, A new method of interpolation and smooth curve fitting based on local procedures, J. ACM **17**, 589 (1970).

[184] H. Nakano, T. Narikawa, K. Oohara, K. Sakai, H. Shinkai, H. Takahashi, T. Tanaka, N. Uchikata, S. Yamamoto, and T. S. Yamamoto, Comparison of various methods to extract ringdown frequency from gravitational wave data, Phys. Rev. D **99**, 124032 (2019).

[185] Y. Yuan, X.-L. Fan, H.-J. Lv, Y.-Y. Sun, and K. Lin, Waveform reconstruction of core-collapse supernovae gravitational-waves with ensemble empirical mode decomposition, arXiv:2309.06011.

[186] H. A. Bethe and J. R. Wilson, Revival of a stalled supernova shock by neutrino heating, Astrophys. J. **295**, 14 (1985).

[187] J. M. LeBlanc and J. R. Wilson, A numerical example of the collapse of a rotating magnetized star, Astrophys. J. **161**, 541 (1970).

[188] M. J. Szczepańczyk *et al.*, Detecting and reconstructing gravitational waves from the next galactic core-collapse supernova in the advanced detector era, Phys. Rev. D **104**, 102002 (2021).

[189] D. George and E. Huerta, Deep learning for real-time gravitational wave detection and parameter estimation: Results with Advanced LIGO data, Phys. Lett. B **778**, 64 (2018).

[190] S. Hochreiter and J. Schmidhuber, Long short-term memory, Neural Comput. **9**, 1735 (1997).

[191] V. Jahmunah, E. Ng, R.-S. Tan, S. L. Oh, and U. R. Acharya, Explainable detection of myocardial infarction using deep learning models with Grad-CAM technique on ECG signals, Comput. Biol. Med. **146**, 105550 (2022).

[192] H. Panwar, P. Gupta, M. K. Siddiqui, R. Morales-Menendez, P. Bhardwaj, and V. Singh, A deep learning and Grad-CAM based color visualization approach for fast detection of COVID-19 cases using chest X-ray and CT-Scan images, Chaos Solitons Fractals **140**, 110190 (2020).

[193] T. C. Abbott, E. Buffaz, N. Vieira, M. Cabero, D. Haggard, A. Mahabal, and J. McIver, GWSkyNet-Multi: A machine-learning multiclass classifier for LIGO-Virgo public alerts, Astrophys. J. **927**, 232 (2022).

[194] J. Powell and B. Müller, Gravitational wave emission from 3D explosion models of core-collapse supernovae with low and normal explosion energies, Mon. Not. R. Astron. Soc. **487**, 1178 (2019).

[195] D. Radice, V. Morozova, A. Burrows, D. Vartanyan, and H. Nagakura, Characterizing the gravitational wave signal from core-collapse supernovae, Astrophys. J. Lett. **876**, L9 (2019).

[196] J. Powell and B. Müller, Three-dimensional core-collapse supernova simulations of massive and rotating progenitors, Mon. Not. R. Astron. Soc. **494**, 4665 (2020).

[197] J. Powell, B. Müller, and A. Heger, The final core collapse of pulsational pair instability supernovae, Mon. Not. R. Astron. Soc. **503**, 2108 (2021).

[198] B. Müller, H.-T. Janka, and H. Dimmelmeier, A new multi-dimensional general relativistic neutrino hydrodynamic code for core-collapse supernovae. I. Method and code test in spherical symmetry, Astrophys. J. Suppl. Ser. **189**, 104 (2010).

[199] M. A. Skinner, J. C. Dolence, A. Burrows, D. Radice, and D. Vartanyan, Fornax: A flexible code for multiphysics astrophysical simulations, Astrophys. J. Suppl. Ser. **241**, 7 (2019).

[200] J. M. Lattimer and F. D. Swesty, A generalized equation of state for hot, dense matter, Nucl. Phys. A **535**, 331 (1991).

[201] A. W. Steiner, M. Hempel, and T. Fischer, Core-collapse supernova equations of state based on neutron star observations, Astrophys. J. **774**, 17 (2013).

[202] A. Nitz *et al.*, gwastro/pycbc: v2.2.2 release of pycbc.

[203] R. Abbott *et al.* (The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration), Open data from the third observing run of LIGO, Virgo, KAGRA, and GEO, Astrophys. J. Suppl. Ser. **267**, 29 (2023).

[204] P. Welch, The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, IEEE Trans. Audio Electroacoust. **15**, 70 (1967).

[205] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, Curriculum Learning, in *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09 (Association for Computing Machinery, New York, NY, USA, 2009) pp. 41–48.

[206] L. van der Maaten and G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. **9**, 2579 (2008).

[207] B. Abbott, R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. Adhikari, V. Adya, C. Affeldt, *et al.*, GW190425: Observation of a compact binary coalescence with total mass $\sim 3.4\ M_\odot$, Astrophys. J. **892**, L3 (2020).

[208] A. Kolmus, G. Baltus, J. Janquart, T. van Laarhoven, S. Caudill, and T. Heskes, Fast sky localization of gravitational waves using deep learning seeded importance sampling, Phys. Rev. D **106**, 023032 (2022).

[209] Y. Sakai *et al.*, Unsupervised learning architecture for classifying the transient noise of interferometric gravitational-wave detectors, Sci. Rep. **12**, 9935 (2022).

[210] X. Fan, J. Li, X. Li, Y. Zhong, and J. Cao, Applying deep neural networks to the detection and space parameter estimation of compact binary coalescence with a network of gravitational wave detectors, Sci. China Phys. Mech. Astron. **62**, 969512 (2019).

[211] W. Wei, A. Khan, E. Huerta, X. Huang, and M. Tian, Deep learning ensemble for real-time gravitational wave detection of spinning binary black hole mergers, Phys. Lett. B **812**, 136029 (2021).

[212] H. Xia, L. Shao, J. Zhao, and Z. Cao, Improved deep learning techniques in gravitational-wave data analysis, Phys. Rev. D **103**, 024040 (2021).

[213] M. D. Morales, J. M. Antelis, C. Moreno, and A. I. Nesterov, Deep learning for gravitational-wave data analysis: A resampling white-box approach, Sensors **21**, 3174 (2021).

[214] N. Lopac, F. Hržić, I. P. Vuksanović, and J. Lerga, Detection of non-stationary GW signals in high noise from Cohen's class of time-frequency representations using deep learning, IEEE Access **10**, 2408 (2022).

[215] M. B. Schäfer, O. c. v. Zelenka, A. H. Nitz, F. Ohme, and B. Brügmann, Training strategies for deep learning gravitational-wave searches, Phys. Rev. D **105**, 043002 (2022).

[216] M. B. Schäfer and A. H. Nitz, From one to many: A deep learning coincident gravitational-wave search, Phys. Rev. D **105**, 043003 (2022).

[217] C. Ma, W. Wang, H. Wang, and Z. Cao, Ensemble of deep convolutional neural networks for real-time gravitational wave signal recognition, Phys. Rev. D **105**, 083013 (2022).

[218] W.-H. Ruan, H. Wang, C. Liu, and Z.-K. Guo, Rapid search for massive black hole binary coalescences using deep learning, Phys. Lett. B **841**, 137904 (2023).

[219] C. Murali and D. Lumley, Detecting and denoising gravitational wave signals from binary black holes using deep learning, Phys. Rev. D **108**, 043024 (2023).

[220] M. B. Schäfer, F. Ohme, and A. H. Nitz, Detection of gravitational-wave signals from binary neutron star mergers using machine learning, Phys. Rev. D **102**, 063015 (2020).

[221] Y.-C. Lin and J.-H. P. Wu, Detection of gravitational waves using Bayesian neural networks, Phys. Rev. D **103**, 063034 (2021).

[222] G. Baltus, J. Janquart, M. Lopez, H. Narola, and J.-R. Cudell, Convolutional neural network for gravitational-wave early alert: Going down in frequency, Phys. Rev. D **106**, 042002 (2022).

[223] A. Menéndez-Vázquez, M. Kolstein, M. Martínez, and L. M. Mir, Searches for compact binary coalescence events using neural networks in the LIGO/Virgo second observation period, Phys. Rev. D **103**, 062004 (2021).

[224] H. Yu, R. X. Adhikari, R. Magee, S. Sachdev, and Y. 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**, 062004 (2021).

[225] S. Garg, S. Sasaoka, D. S. Dominguez, Y. Hou, N. Koyama, K. Somiya, H. Takahashi, and M. Ohashi, Comparison of training methods for convolutional neural network model for gravitational-wave detection from neutron star-black hole binaries, in *Proceedings of 38th International Cosmic Ray Conference (ICRC2023)* (Sissa Medialab, Nagoya, Japan, 2023).

[226] LIGO Scientific Collaboration, LIGO Algorithm Library - LALSuite, free software (GPL).

[227] R. Sturani, S. Fischetti, L. Cadonati, G. M. Guidi, J. Healy, D. Shoemaker, and A. Viceré, Complete phenomenological gravitational waveforms from spinning coalescing binaries, J. Phys. Conf. Ser. **243**, 012007 (2010).

[228] N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, and A. Iosifidis, Deep adaptive input normalization for time series forecasting, IEEE Trans. Neural Netw. Learn. Syst. **31**, 3760 (2020).

[229] A. L. Maas, A. Y. Hannun, A. Y. Ng, *et al.*, Rectifier nonlinearities improve neural network acoustic models, in *Proceedings of the ICML Workshop on Deep Learning for Audio, Speech and Language Processing* (ICML, Atlanta, GA, USA, 2013).

[230] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, arXiv:1207.0580.

[231] A. Paszke *et al.*, PyTorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems 32* (Curran Associates, Vancouver, Canada, 2019) pp. 8024–8035.

[232] N. Kokhlikyan *et al.*, Captum: A unified and generic model interpretability library for PyTorch, arXiv:2009.07896.

[233] A. Mezzacappa and M. Zanolin, Gravitational waves from neutrino-driven core collapse supernovae: Predictions, detection, and parameter estimation, arXiv:2401.11635.

[234] A. Torres-Forné, P. Cerdá-Durán, M. Obergaulinger, B. Müller, and J. A. Font, Universal relations for gravitational-wave asteroseismology of protoneutron stars, Phys. Rev. Lett. **123**, 051102 (2019).

[235] A. Torres-Forné, P. Cerdá-Durán, M. Obergaulinger, B. Müller, and J. A. Font, Erratum: Universal relations for gravitational-wave asteroseismology of protoneutron stars [Phys. Rev. Lett. 123, 051102 (2019)], Phys. Rev. Lett. **127**, 239901 (2021).

[236] H. Sotani, T. Takiwaki, and H. Togashi, Universal relation for supernova gravitational waves, Phys. Rev. D **104**, 123009 (2021).

[237] M.-A. Bizouard, P. Maturana-Russel, A. Torres-Forné, M. Obergaulinger, P. Cerdá-Durán, N. Christensen, J. A. Font, and R. Meyer, Inference of protoneutron star properties from gravitational-wave data in core-collapse supernovae, Phys. Rev. D **103**, 063006 (2021).

[238] J. Powell and B. Müller, Inferring astrophysical parameters of core-collapse supernovae from their gravitational-wave emission, Phys. Rev. D **105**, 063018 (2022).

[239] T. M. Tauris, N. Langer, and P. Podsiadlowski, Ultra-stripped supernovae: Progenitors and fate, Mon. Not. R. Astron. Soc. **451**, 2123 (2015).

[240] *Updated Advanced LIGO sensitivity design curve*, Tech. Rep. T1800044-v5 (LIGO Document, 2018).

[241] S. Hild *et al.*, Sensitivity studies for third-generation gravitational wave observatories, Classical Quantum Gravity **28**, 094013 (2011).

[242] M. Drago *et al.*, Coherent WaveBurst, A pipeline for unmodeled gravitational-wave data analysis, SoftwareX **14**, 100678 (2021).

[243] V. Necula, S. Klimenko, and G. Mitselmakher, Transient analysis with fast Wilson-Daubechies time-frequency transform, J. Phys. Conf. Ser. **363**, 012032 (2012).

[244] S. Klimenko, I. Yakushin, A. Mercer, and G. Mitselmakher, A coherent method for detection of gravitational wave bursts, Classical Quantum Gravity **25**, 114029 (2008).

[245] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration and ASAS-SN Collaboration and DLT40 Collaboration), Optically targeted search for gravitational waves emitted by core-collapse supernovae during the first and second observing runs of Advanced LIGO and Advanced Virgo, Phys. Rev. D **101**, 084002 (2020).

[246] B. P. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Observing gravitational-wave transient GW150914 with minimal assumptions, Phys. Rev. D **93**, 122004 (2016).

[247] M. Lopez, V. Boudart, K. Buijsman, A. Reza, and S. Caudill, Simulating transient noise bursts in LIGO with generative adversarial networks, Phys. Rev. D **106**, 023027 (2022).

[248] J. Yan, A. P. Leung, and C. Y. Hui, On improving the performance of glitch classification for gravitational wave detection by using generative adversarial networks, Mon. Not. R. Astron. Soc. **515**, 4606 (2022).

[249] N. J. Cornish and T. B. Littenberg, Bayeswave: Bayesian inference for gravitational wave bursts and instrument glitches, Classical Quantum Gravity **32**, 135012 (2015).

[250] T. B. Littenberg and N. J. Cornish, Bayesian inference for spectral estimation of gravitational wave detector noise, Phys. Rev. D **91**, 084034 (2015).

[251] N. J. Cornish, T. B. Littenberg, B. Bécsy, K. Chatziioannou, J. A. Clark, S. Ghonge, and M. Millhouse, BayesWave analysis pipeline in the era of gravitational wave observations, Phys. Rev. D **103**, 044006 (2021).

[252] R. Demirli and J. Saniie, Asymmetric Gaussian chirplet model and parameter estimation for generalized echo representation, J. Franklin Inst. **351**, 907 (2014).

[253] V. Roma, J. Powell, I. S. Heng, and R. Frey, Astrophysics with core-collapse supernova gravitational wave signals in the next generation of gravitational wave detectors, Phys. Rev. D **99**, 063018 (2019).

[254] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (W.H. Freeman, San Francisco, 1973).

[255] L. Cohen, The uncertainty principle in signal analysis, in *Proceedings of IEEE-SP International Symposium on Time- Frequency and Time-Scale Analysis* (IEEE, Philadelphia, PA, USA, 1994) pp. 182–185.

[256] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide*, 3rd ed. (Society for Industrial and Applied Mathematics, Philadelphia, PA, 1999).