Continuous Waves detection based on a Transfer Learning approach

by

Diego Domínguez

ORCID: 0009-0007-7550-1933

A thesis submitted in total fulfillment for the
Master's degree in Physics

in the
Department of Physics
School of Science
TOKYO INSTITUTE OF TECHNOLOGY

July 2024

TOKYO INSTITUTE OF TECHNOLOGY

Abstract

Department of Physics
School of Science

Master in Physics

by Diego Domínguez
ORCID: 0009-0007-7550-1933

Gravitational waves (GWs) have been a longstanding prediction of Einstein's General Relativity for over a century. The era of GW astrophysics was inaugurated in 2015 by the detection of a signal from a binary black hole merger. Since then, significant progress has been made in the field, making signal detection routine. This advancement has revolutionized astrophysics, cosmology, and our understanding of the universe and our galactic vicinity. However, much work remains to be done. The detection of GWs has opened new avenues for studying the Universe and testing fundamental physics. As LIGO, Virgo and KAGRA undertake another observation run (O4) with an improved sensitivity, non-axisymmetric neutron stars emitting quasi monochromatic, long-standing GWs are expected to be within the detectors' sensitivity band. However, their detection in the presence of noise remains a challenging problem. In recent years, Deep Learning approaches have been proposed as a potential solution to this issue. This study explores the effectiveness of employing a transfer-learning approach with a binary-classifier network, based on ResNext50, for detecting unmodeled CWs embedded in Gaussian and real-detector noise from the O3 observation run. This model leverages spectrograms derived from the H1 and L1 detectors, divided into numerous segments each extending over approximately four-days periods. Such method is compared to the SOAP CW pipeline, showing improved efficiency and generalization to lower signal-to-noise ratios. Through this technique, inference is applied to several data segments to accumulate a comprehensive detection statistic across frequency bins. The results show the benefits of the use of more complex network-architectures and set a precedent for the further development of the detection of unmodeled CW signals based on artificial neural networks. Although progress has been made, further work is needed regarding the choice of a semicoherent scheme adapted to longer observation times. Several options for future research on the studied CNN architecture are proposed.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Kentaro Somiya. His invaluable guidance and support over the past two years have been instrumental in my academic journey. His profound insights and extensive experience in the field of Gravitational Wave Physics have been a source of inspiration for me as an emerging scientist. I am particularly grateful for the numerous opportunities he provided for me to present my work at various conferences and workshops, which have significantly contributed to my professional growth.

I also wish to thank Hirotaka Takahashi for his great support and guidance during our meetings and for providing access to the computational resources essential for completing this work. I am deeply grateful to the members of the LIGO-Virgo-KAGRA collaboration, especially the Continuous Waves working group, for their invaluable feedback on my presentations. I extend my heartfelt appreciation to all the members of the Continuous Waves community I met in Hanover. Their input has profoundly shaped my understanding of the current developments in the field and has enriched my knowledge, inspiring new scientific ideas and research avenues.

My sincere thanks go to all the members of the Somiya Lab, with whom I had the privilege of working daily. Their collaborative spirit and willingness to share ideas have greatly enhanced my academic experience. Special thanks to Seiya Sasaoka, whose insightful discussions and technical knowledge have significantly influenced the direction of my research and improved my coding skills. His dedication and hard work have been a constant source of inspiration. I am also grateful to Kaido Suzuki for his experience and insights into Japanese culture, which have been incredibly helpful. His friendship and reliability in the lab have been invaluable. I extend my thanks to all the members who have come and gone; each one has contributed to a balanced and enriching academic experience at Tokyo Tech.

I am deeply appreciative of the many friends and brilliant individuals I have met in Japan. Special thanks to my dear friends Noor and Ayberk, whose friendship has been crucial to my life here. Their skills, humor, and camaraderie have been a constant source of amazement and support. It truly would not have been the same without them. My deepest gratitude goes to Hira, who has tirelessly supported me with much patience and kindness.

None of my achievements would have been possible without the unwavering support of my family and friends back in Ecuador. Their sacrifice, hard work, and relentless support throughout the years have made everything possible. Lastly, I would like to thank the people of Japan and its government for granting me the opportunity to pursue my graduate studies through the MEXT scholarship and for allowing me to explore and appreciate the wonderful land of Japan.

Contents

List of Figures

List of Tables

Abbreviations

AXP	**A**nomalous **X**-ray **P**ulsar
BBH	**B**inary **B**lack **H**ole
BBN	**B**ig **B**ang **N**ucleosynthesis
BSD	**B**and-**S**ampled **D**ata
CBC	**C**ompact **B**inary **C**oalescence
CCO	**C**entral **C**ompact **O**bject
CMBR	**C**osmic **M**icrowave **B**ackground **R**adiation
CNN	**C**onvolutional **N**eural **N**etwork
CW	**C**ontinuous (Gravitational) **W**ave
DECIGO	**DEC**i-Hertz **I**nterferometer **G**ravitational Wave **O**bservatory
DFT	**D**iscrete **F**ourier **T**ransform
DL	**D**eep **L**earning
DNN	**D**eep **N**eural **N**etwork
EFE	**E**instein **F**ield **E**quations
EM	**E**lectro **M**agnetic
ET	**E**instein **T**elescope
FFT	**F**ast **F**ourier **T**ransform
FP	**F**abry-**P**erot
GDM	**G**radient **D**escent **M**omentum
GPU	**G**raphics **P**rocessing **U**nit
GR	**G**eneral **R**elativity
GW	**G**ravitational **W**ave
HMM	**H**idden **M**arkov **M**odel
ISM	**I**nter-**S**tellar **M**edium
LAL	**L**IGO **A**lgorithm **L**ibrary
LIGO	**L**aser **I**nterferometer **G**ravitational Wave **O**bservatory
LISA	**L**aser **I**nterferometer **S**pace **A**ntenna
LMXB	**L**ow **M**ass **X**-ray **B**inary
MCMC	**M**arkov **C**hain **M**onte **C**arlo
NAG	**N**esterov **A**ccelerated **G**radient
NS	**N**eutron **S**tar

PN	**P**ost **N**ewtonian
QCD	**Q**uantum **C**hromo **D**ynamics
RMS	**R**oot **M**ean **S**quare
RNN	**R**ecurrent **N**eural **N**etwork
SFDB	**S**hort **F**FT **D**ata **B**ase
SFT	**S**hort **F**ourier **T**ransform
SGD	**S**tochastic **D**escent **G**radient
SGR	**S**oft **G**amma **R**ay
SMBH	**S**uper **M**assive **B**lack **H**ole
SMBHM	**S**uper **M**assive **B**lack **H**ole **M**erger
SNR	**S**ignal-to-**N**oise **R**atio
SSB	**S**olar **S**ystem **B**arycenter
TT	**T**ransverse **T**raceless

Notation and conventions

Indices, metric signature, four-vectors

Greek Indices like $\alpha, \beta, \cdots$ or $\mu, \nu, \cdots$ can take the values $0, \cdots, 3$ while spatial indices are denoted by latin letters $i, j, \cdots$ and they take the values $1, 2, 3$. The flat-space (Minkowski) metric is $\eta_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$, which is the usual signature in General relativity. We also consider

$$\begin{aligned} x^\mu &= (x^0, \boldsymbol{x}), \qquad x^0 = ct, \\ \partial_\mu &= \frac{\partial}{\partial x^\mu} = \left(\frac{1}{c}\partial_t, \partial_i\right), \\ \Box &= \eta_{\mu\nu}\partial^\mu\partial^\nu \\ d^4x &= dx^0\ d^3x = cdt\ d^3x \end{aligned}$$

A dot denotes the time derivative $\dot{f}(t) = \partial_t f = c\partial_0 f$. Following the Einstein summation convention, repeated upper and lower indices are summed over.

Important tensors in General Relativity

The definitions of the Christoffel symbols, Riemann tensor, Ricci tensors and scalar curvature are as follows

$$\begin{aligned} \Gamma^\rho{}_{\mu\nu} &= \frac{1}{2}g^{\rho\sigma}\left(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}\right), \\ R^\mu{}_{\nu\rho\sigma} &= \partial_\rho\Gamma^\mu{}_{\nu\sigma} - \partial_\sigma\Gamma^\mu{}_{\nu\rho} + \Gamma^\mu{}_{\alpha\rho}\Gamma^\alpha{}_{\nu\sigma} - \Gamma^\mu{}_{\alpha\sigma}\Gamma^\alpha{}_{\nu\rho}, \\ R_{\mu\nu} &= R^\alpha{}_{\mu\alpha\nu} \\ R &= g^{\mu\nu}R_{\mu\nu} \end{aligned}$$

Introduction

Gravitational waves (GWs) are a direct consequence of Einstein's general theory of relativity, first predicted in 1915 as ripples in space-time that propagate at the speed of light. Initially, they were considered nonphysical and merely mathematical curiosities that could be "gauged" away by coordinate transformation. However, the existence of GWs is now firmly established. Over the past century, numerous theoretical and experimental advancements have been made. Significant milestones include Joseph Weber's antenna [1] in 1958, the observational evidence from the Hulse-Taylor pulsar [2] in 1974, and the first direct detection of GWs [3] by the Laser Interferometer Gravitational-Wave Observatory (LIGO) in 2015, which identified a signal from a binary black hole (BBH) merger.

The study of GWs has become a highly active area of research, with significant advancements both experimentally and theoretically. After decades of development, GW detectors have achieved the sensitivity needed to detect hundreds of events each year, making observations of Compact Binary Coalescences (CBC) involving black holes and neutron stars routine. A notable breakthrough in this field is the development of multimessenger astronomy using a gravitational channel, exemplified by the GW170817 event [4]. This event, an inspiral of a binary neutron star system observed through both gravitational and electromagnetic signals, provided crucial insights, such as constraints on the speed of gravity relative to light, and challenged some modified gravity theories [5]. The integration of gravitational and electromagnetic observations is just one example of the many possible new research opportunities that the field has to offer, setting the stage for further discoveries with the upcoming deployment of next-generation detectors like LISA, DECIGO, Cosmic Explorer, and the Einstein Telescope.

Parallel to these experimental advancements, theoretical research has also made significant progress. Our understanding of the mechanisms generating GWs, spanning both astrophysical and cosmological contexts, has deepened. Detecting long-lasting Continuous Waves, transient signals, the stochastic background, and GWs from events such as stellar collapses and supermassive black hole activity, as well as hypothetical sources like dark matter or cosmic strings, is becoming increasingly feasible. Therefore, it is crucial to prepare for possible detection scenarios, especially for signals that are theoretically within the sensitivity band of current detectors but are challenging to analyze due to the computational expense of matched-filtering algorithms. This thesis focuses on the continuous gravitational radiation emitted by rapidly spinning neutron stars, known as *Continuous Gravitational Waves* (CWs), and explores their detection using a *Deep Learning* (DL) based approach. To provide the necessary background, Chapter 1 introduces the fundamental concepts of General Relativity and the framework of gravitational wave

physics. Following this, Chapter 2 delves into the specifics of CWs and their detection. The foundations of DL techniques follow in chapter 3. The application of DL techniques to this detection problem and the main results of this study and discussion are presented in Chapter 4.

Chapter 1

Gravitational Waves

This chapter introduces the principles of GW physics as described by GR and provides an overview of the detectors and astrophysical sources. While a detailed derivation of the main results from GR is not presented, this chapter offers a concise and accessible introduction, serving as a roadmap to the fundamental concepts necessary for understanding the main research themes discussed in the subsequent chapters.

1.1 General Relativity

General Relativity (GR) concerns the geometrical description of the curvature of space-time caused by the presence of matter and energy. Also known as "Geometrodynamics," the central idea around the equations governing the gravitational interaction is that matter tells space how to curve, and space tells matter how to move. From the principle of least action, one can write the action as an integral of a Lagrangian density taken over space-time, as the sum of a contribution of the field $\mathcal{L}_E$ and a contribution from matter $\mathcal{L}_M$

$$S = \int d^4x \ \ (\mathcal{L}_E + \mathcal{L}_M)\,. \tag{1.1}$$

Following a similar approach to Newtonian gravity, one might incorporate into the field part of the action a quantity that depends on the second derivatives of the coordinates. In a purely geometrical description, the only independent scalar that can be constructed from the metric tensor $g_{\mu\nu}$ and that contains second partial derivatives is the scalar curvature R. Hilbert demonstrated that the simplest form of this is

$$\mathcal{L}_E = \frac{c^3}{16\pi G}\sqrt{-g}R, \tag{1.2}$$

where G is the gravitational constant and c is the speed of light. The equations of motion follow from the variation of the action with respect to the metric $g_{\mu\nu} \to g_{\mu\nu} + \delta g_{\mu\nu}$

$$\delta S_E = \int d^4x \left(\sqrt{-g}g^{\mu\nu}\delta R_{\mu\nu} + \sqrt{-g}R_{\mu\nu}\delta g^{\mu\nu} + R\delta\sqrt{-g}\right), \tag{1.3}$$

where the scalar curvature is obtained from the contraction of the Ricci tensor $R = g^{\mu\nu}R_{\mu\nu}$. Computing and grouping terms with respect to the variation of the metric, one obtains

$$\delta S_E = \int d^4x\sqrt{-g}\left(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}\right)\delta g^{\mu\nu} \tag{1.4}$$

On the other hand, the matter component of the action involves the energy-momentum tensor of matter $T^{\mu\nu}$ which is given by the variation of the metric according to

$$\delta S_M = \frac{1}{2c}\int d^4x\sqrt{-g}T^{\mu\nu}\delta g_{\mu\nu} \tag{1.5}$$

By requiring that $\delta S = \delta S_E + \delta S_M$ vanishes for arbitrary variations, one finds the central equations of GR, the Einstein Field Equations (EFE)

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu} \tag{1.6}$$

1.2 Gravitational Waves

Just as "water waves" can be thought of as small ripples rolling across the ocean, GWs can be identified as tiny ripples propagating through space-time. Mathematically, this identification involves separating the curved space metric into a slowly varying *background* part (the ocean) and a *perturbation* part (the small ripples). In terms of the metric, the EFE can be expanded around a flat-space background as:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad |h_{\mu\nu}| \ll 1, \tag{1.7}$$

where $\eta = \mathrm{diag}(-1, 1, 1, 1)$ is the Minkowski metric and $h_{\mu\nu}$ is the metric perturbation. In this *linearized theory*, the equations of motion are expanded to linear order in $h_{\mu\nu}$. General Relativity is invariant under arbitrary coordinate transformations, a property known as "gauge" invariance. However, for practical experimental measurements, we choose a specific reference frame, thereby breaking this invariance. In such a frame, within the standard Lorentz gauge, the field $h_{\mu\nu}$ must satisfy the conditions

$$h_{\mu\nu} = \bar{h}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\left(\eta^{\rho\sigma}h_{\rho\sigma}\right), \qquad \partial^\nu\bar{h}_{\mu\nu} = 0. \tag{1.8}$$

In this gauge, after a straightforward computation, the equations of motion (1.6) reduce to

$$\Box\bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4}T_{\mu\nu}, \tag{1.9}$$

where the $\Box$ operator is the d'Alembertian $\Box = \eta_{\mu\nu}\partial^\mu\partial^\nu$. For further details on the derivation of Eq. (1.9) see Appendix A.1. Physically, this linearized theory approximation implies that the bodies that act as sources of GWs are assumed to move in flat space-time, along trajectories determined by their mutual gravitational influences. For a self-gravitating system, because the background space-time metric is $\eta_{\mu\nu}$, their dynamics are described by Newtonian gravity rather than full GR. The response of test masses to

the GW part $h_{\mu\nu}$ generated by these bodies is computed using $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ neglecting terms $O(h^2)$ when evaluating $R_{\mu\nu}$. The Lorentz gauge reduces the ten independent components of the metric to six. However, further simplifications can be made to Eq. (1.9). To study the propagation of GWs, we focus on a region of space-time outside the source, i.e. $T^{\mu\nu} = 0$, resulting in

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

which is the well-known wave equation. Additionally, since the Lorentz gauge conditions from Eqs. (1.8) are invariant under the coordinate transformation $x^\mu \to x^\mu + \xi^\mu$, we can impose the *transverse-traceless gauge* (TT gauge) defined by

$$h^{0\mu} = 0, \qquad h^i{}_i = 0, \qquad \partial^j h_{ij} = 0 \tag{1.11}$$

to further reduce the number of independent components of the metric to just two. The first condition ensures that only the spatial components are nonzero, the second guarantees a trace-free metric, and the third imposes that the spatial components are divergence-free. The wave equation has plane wave solutions of the form

$$h_{ij}^{TT} = e_{ij}(\mathbf{k}) e^{i k_a x^a}, \tag{1.12}$$

where $k^\mu = (\omega/c, \mathbf{k})$, $\omega/c = |\mathbf{k}|$ and $e_{ij}(\mathbf{k})$ is the polarization tensor. By choosing $\mathbf{n}$ along the z axis and imposing the TT gauge, the solution to Eq. (1.10) can be written as

$$h_{ij}^{TT}(t,z) = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix} \cos[\omega(t - z/c)] \tag{1.13}$$

where h_+ and $h_\times$ are known as the "plus" and "cross" polarizations of the wave respectively. It's useful to express this result in the plane-wave expansion. The solution for $\Box h_{ij}^{TT} = 0$ can be expressed as

$$h_{ij}^{TT}(x) = \int \frac{d^3k}{(2\pi)^3} \left(\mathcal{A}_{ij}(\boldsymbol{k}) e^{ikx} + \mathcal{A}^*_{ij}(\boldsymbol{k}) e^{-ikx} \right) \tag{1.14}$$

Taking into account that $d^3k = |\boldsymbol{k}|^2 dk\, d\Omega = (2\pi/c)^3 f^2 df\, d\Omega$, the above equation reads

$$h_{ij}^{TT} = \frac{1}{c^3} \int_0^\infty df\, f^2 \int d^2\hat{\boldsymbol{n}} \left(\mathcal{A}_{ij}(f, \hat{\boldsymbol{n}}) e^{-2\pi i f(t - \hat{\boldsymbol{n}}\cdot\boldsymbol{x}/c)} + c.c \right) \tag{1.15}$$

Now, by explicitly defining the polarization tensors as

$$e^+_{ij}(\hat{\boldsymbol{n}}) = \hat{\boldsymbol{u}}_i \hat{\boldsymbol{u}}_j - \hat{\boldsymbol{v}}_i \hat{\boldsymbol{v}}_j, \qquad e^\times_{ij} = \hat{\boldsymbol{u}}_i \hat{\boldsymbol{v}}_j + \hat{\boldsymbol{v}}_i \hat{\boldsymbol{u}}_j \tag{1.16}$$

where $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{v}}$ are unit vectors orthogonal to the propagation direction $\hat{\boldsymbol{n}}$ and to each other. Without loss of generality, a frame where $\hat{\boldsymbol{n}} = \hat{z}$ can be chosen so that

$$e^+_{ab} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}_{ab}, \qquad e^\times_{ab} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}_{ab}, \tag{1.17}$$

Then, in a generic frame, the amplitudes $\tilde{h}_A(f,\hat{\boldsymbol{n}})$ can be defined by

$$\frac{f^2}{c^3}\mathcal{A}_{ij}(f,\hat{\boldsymbol{n}}) = \sum_{A=+,\times} \tilde{h}_A(f,\hat{\boldsymbol{n}}) e^A_{ij}(\hat{\boldsymbol{n}}) \tag{1.18}$$

so that Eq. (1.15) becomes

$$h_{ab}(t,\boldsymbol{x}) = \sum_{A=+,\times} \int_{-\infty}^{\infty} df \int d^2\hat{\boldsymbol{n}}\, \tilde{h}_A(f,\hat{\boldsymbol{n}})\, e^A_{ab}(\hat{\boldsymbol{n}})\, e^{-2\pi i f(t-\hat{\boldsymbol{n}}\cdot\boldsymbol{x}/c)} \tag{1.19}$$

defining implicitly that $\tilde{h}_A(-f,\hat{\boldsymbol{n}}) = \tilde{h}^*_A(f,\hat{\boldsymbol{n}})$. Additionally, to better understand the effects of a GW on test masses, it is important to study the *proper detector frame* using the geodesics equation

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu{}_{\nu\rho}\frac{dx^\nu}{d\tau}\frac{dx^\rho}{d\tau} = 0, \tag{1.20}$$

which is the classical equation of motion of test masses in the curved background described by the metric $g_{\mu\nu}$. When considering two nearby geodesics parametrized by $x^\mu(\tau)$ and $x^\mu(\tau)+\xi^\mu(\tau)$, replacing both in Eq. (1.20) and taking their difference, up to first order in ξ, one gets

$$\frac{d^2\xi^\mu}{d\tau^2} + 2\Gamma^\mu{}_{\nu\rho}(x)\frac{dx^\nu}{d\tau}\frac{d\xi^\rho}{d\tau} + \xi^\sigma\partial_\sigma\Gamma^\mu{}_{\nu\sigma}(x)\frac{dx^\nu}{d\tau}\frac{dx^\rho}{d\tau} = 0, \tag{1.21}$$

which is the equation for geodesics deviation. Assuming the detector moves non-relativistically, $dx^i/d\tau$ can be neglected compared to $dx^0/d\tau$. Given that $\Gamma^\mu{}_{\nu\rho}$ vanishes at the expansion point P and that $\partial_\sigma\Gamma^i{}_{00}$ is evaluated at the point P, this simplifies to

$$\frac{d^2\xi^i}{d\tau^2} = -R^i{}_{0j0}\left(\frac{dx^0}{d\tau}\right)^2. \tag{1.22}$$

In a linearized theory, the Riemann tensor is invariant, so we compute $R^i{}_{0j0}$ in the TT frame, finding that the geodesics deviation in the proper detector frame up to linear order in h, where $t=\tau$, is

$$\ddot{\xi}^i = \frac{1}{2}\ddot{h}^{TT}_{ij}\xi^j, \tag{1.23}$$

where the dot denotes the derivative with respect to the coordinate time t. One remarkable consequence of this expression is that the effects of the GW in the detector frame can be described by a *Newtonian*-like force without further reference to GR. Thus, in practice, as long as we can sufficiently isolate the detector from the influence of other slowly varying Newtonian forces (implicit in the approximations made for the Christoffel symbols and their derivatives), the effect on test masses can be measured in a frequency window $[f_{\min}, f_{\max}]$ where $f_{\min}$ is sufficiently large to control slowly varying Newtonian noises. Neglecting terms beyond linear order in ξ defines a characteristic scale over which the gravitational field remains approximately constant. For GW, this length scale is the reduced wavelength $\lambda\!\!\!^{-}$. Hence, if a detector has a linear size L, Eq. (1.23) is valid if and only if $L \ll \lambda\!\!\!^{-}$.

1.2.1 Fermi coordinates

As mentioned before, it's possible to give a quasi-Newtonian description of the motion of point masses under the action of GWs. To get even more insight, we can use the *Fermi coordinates*, i.e., a local inertial frame defined in a neighborhood of an entire timelike geodesic, that deviates from the flat metric only quadratically in the distance from this world-line. Thus, using Fermi coordinates x^μ defined in the vicinity of the world-line of a freely falling observer , the space-time interval takes the form

$$ds^2 = -(dx^0)^2 + \delta_{ij}dx^i dx^j + \mathcal{O}(|x^i|^2/d^2), \tag{1.24}$$

where deviations from the flat space metric appear at a quadratic order in the spatial distance $|x^i|^2$ and d is the curvature radius such that $|R_{\alpha\beta\mu\nu}| \sim d^{-2}$. The coordinate transformation from the TT gauge to the Fermi gauge given by the last equation [6] reads

$$\hat{x}^0 = t, \qquad \hat{x}^i = x^i + \frac{1}{2}h_{ij}^{TT}(t,\vec{0})x^j \tag{1.25}$$

where $h_{ij}^{TT}(t,\vec{0})$ denotes the value of the field h_{ij}^{TT} along the world-line of the free-falling observer. In the particular case of a monochromatic wave of frequency ω (as measured by the observer) that propagates along the direction $\hat{z}$ direction. Using $h_{+,\times}(t) = H_{+,\times}e^{i\omega t}$ gives

$$\hat{x}(t) = x_0 + \frac{1}{2}\left(H_+ x_0 + H_\times y_0\right)e^{i\omega t} \tag{1.26}$$

$$\hat{y}(t) = y_0 + \frac{1}{2}\left(H_\times x_0 + H_+ y_0\right)e^{i\omega t} \tag{1.27}$$

$$\hat{z}(t) = z_0 \tag{1.28}$$

Thus, a GW that propagates through an initial circular ring of particles induces alternative contractions and elongations along the $\hat{x}$ and $\hat{y}$ directions for the + polarization, and along the $\hat{y} = \hat{x}$ and $\hat{y} = -\hat{x}$ directions for the $\times$ polarizations (See Fig. 1.1). A generic GW can thus be understood as a superposition of two oscillating tidal fields that propagate at the vacuum speed of light.

1.2.2 The Quadrupolar nature of GWs

Without the aid of GR, it is possible to get an initial insight about the nature of GW radiation. From electromagnetism, we know that the electric- and magnetic-dipole radiation terms (proportional to the second derivatives of the dipole moment $\ddot{d}$ and the magnetic moment $\ddot{\mu}$) dominate the power output. For the gravitational case, to get a rough approximation, with the substitution $e^2 \to -m^2$ we can approximate Newton's law of attraction with Coulomb law. In this case by using the gravitational analog of the dipole moments

$$\boldsymbol{d} = \sum(\text{“charge”})(\text{position}) = \sum m_i \mathbf{x}_i = \mathbf{p}, \tag{1.29}$$

$$\boldsymbol{\mu} = \sum(\text{position}) \times (\text{“current”}) = \sum \mathbf{r}_i \times (m\mathbf{v}_i) = \mathbf{J}, \tag{1.30}$$

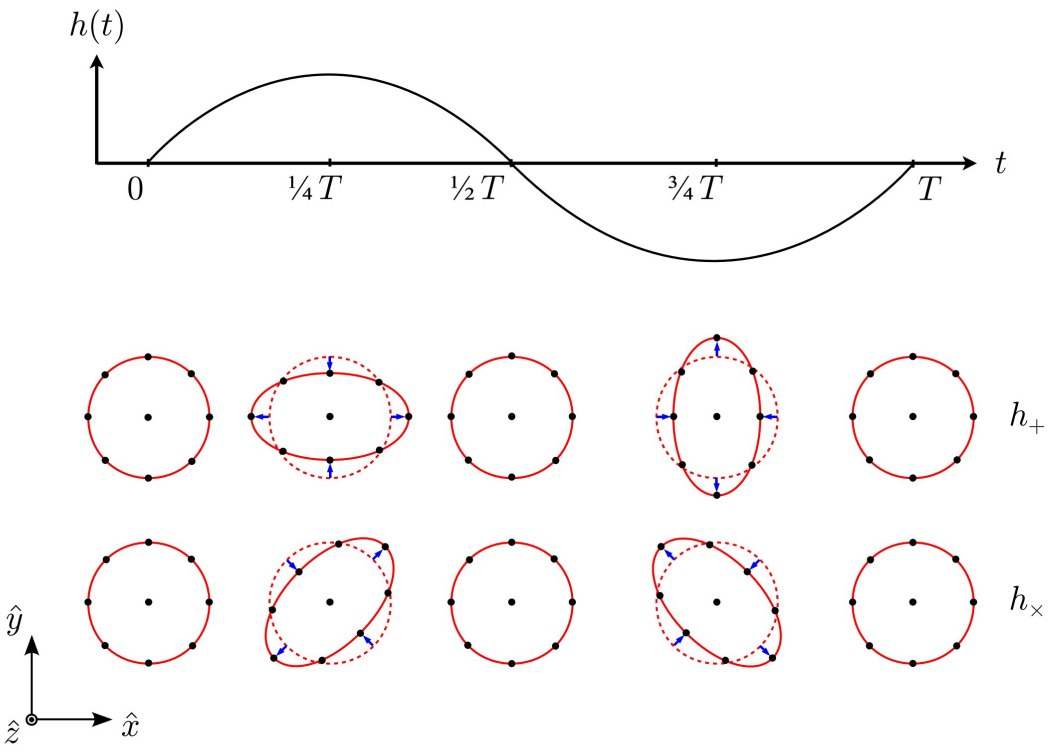


FIGURE 1.1: A monochromatic GW of frequency ω propagates along the $\hat{z}$ direction. The lower panel shows the effects of the + and the × polarizations on a ring of freely falling particles in a local inertial frame.

it is immediately clear that by virtue of the conservation of linear and angular momentum, there can be no mass dipole radiation of any sort. However, following the derivation found in [7], in the slow-motion expansion one can solve Eq. (1.9) in terms of the second moment of the mass distribution Q_{ij} as

$$h_{ij}^{TT}(t,r) = \frac{2G}{rc^4}\frac{d^2}{dt^2}Q_{ij}^{TT}(t-r/c), \tag{1.31}$$

where r is the distance from the source and the second moment of the mass distribution is given by

$$Q_{ij}(t) = \int \rho(t,\boldsymbol{x})\; x^i x^j \;\; d^3x. \tag{1.32}$$

Here $\rho = T^{00}/c^2$ is the mass density. The energy flow per unit time through a unit surface at a large distance from the source is given by

$$\frac{dE}{dt} = \frac{c^3 r^2}{32\pi G}\int d\Omega\; \langle \dot{h}_{ij}^{TT}\dot{h}_{ij}^{TT}\rangle \tag{1.33}$$

Consequently, by integrating over the solid angle and taking the temporal average over several characteristic periods of the GW, one finds that the total radiated power is

$$P = \frac{G}{5c^5}\langle \dddot{Q}_{ij}\,\dddot{Q}_{ij}\rangle, \tag{1.34}$$

making manifest that the leading term of GW radiation is quadrupolar in nature. The fact that electromagnetic radiation is predominantly dipolar and gravitational is quadrupolar are consequences of a

general theorem [7]. Consider a classical field, whose associated quantum particles have integer spin S and zero rest mass. Expand the radiation field into spherical harmonics (multipole moments). All components with $l < S$ will vanish and this is independent of the nature of the source. Gravitational radiation possesses a $S = 2$, hence its leading term for slowly moving sources $v \ll c$ should be quadrupolar.

1.3 Detectors of GWs

Designing GW detectors is analogous to designing radio receivers, except that electric charges moving freely in conductors are replaced by test masses. This concept is illustrated in Fig. 1.1, showing how a ring of particles is deformed by a passing GW. The first of such detectors was constructed by Weber [1] in the 1960s. It was a large stiff bar in which GWs could induce quadrupole vibrations. Unfortunately, as it was later shown, its sensitivity was at least ten orders of magnitude below what was required to make a confident detection. Carrying out on this line of work, the International GW Event Collaboration (IGEC) used five cryogenic resonant mass detectors: ALLEGRO [8], AURIGA [9], EXPLORER and NAUTILUS [10] and NIOBE [11]. Later, instruments like MiniGRAIL [12] and Schenberg [13] were designed using spherical antennas with the hope of determining both source direction and wave polarisation. Today, most research is focused on detectors that consist of a distribution of at least two test masses that are coupled together by an electromagnetic field (such as a laser beam) instead of mechanical forces. In almost all cases, an elaborate Michelson interferometer is the standard choice because of its ability to measure relative displacements in orthogonal directions, whilst suppressing the noise of the light source used for the measurement.

Depending on the frequency band of interest, different factors dominate the design of detectors. At lower frequencies, the *nanohertz* band (fraction of a cycle per year), the *pulsar timing* technique uses pairs of test masses consisting of the Earth and certain distant milisecond pulsars. While quite tiny, the EM field of the pulsar beam provides the coupling field of detection. Detection depends on the precise timing between different Earth-pulsar pairs. The International Pulsar Timing Array project and the NANOGrav collaboration carry on detection using this technique and have recently found evidence [14] for the GW Background of the universe. At around one cycle per hour, *space laser interferometry* has been proposed. This technique requires sets of kg-scale tests masses linked by low-power laser beams on a multi-million km baseline. The Laser Interferometer Space Antenna (LISA) [15] consists of three spacecraft placed at the vertices of an equilateral triangle of size 5×10^6 km and set in solar orbit at 1 AU 20^o behind the Earth. At around 0.1 - 1 Hz another space laser interferometer with 1000 km baselines has been proposed. DECIGO [16], the Deci-hertz Interferometer Gravitational Wave Observatory, consisting of three drag-free satellites 1000 km apart. The relative displacements between the craft would be measured using three sets of Fabry-Perot Michelson Interferometers.

In the terrestrial *audio frequency* band (about 10 Hz to $\sim$ 3 kHz), changes in the spacing of test masses kilometers apart are measured with very high power laser interferometers. These include LIGO, Virgo, GEO600 and KAGRA. Third generation detectors are also expected to be added to the network in the

TABLE 1.1: A comparison of various detectors of GWs

Frequency band	Technique	Projects
Nanohertz	Pulsar timing	NANOGrav [17] Parkes Pulsar Timing array [18] International Pulsar Timing Array [19] European Pulsar Timing Array [20]
Milihertz	Space transponder interferometers	Laser Interferometer Space Antenna (LISA) [21]
Decihertz	Fabry-Perot space laser interferometer	Deci-hertz Interferometric Gravitational Wave Observatory [16]
Audio	Terrestrial laser interferometers	Advanced LIGO [22] Advanced Virgo [23] KAGRA [24] TAMA [25] LCGT [26] GEO [27] AIGO [28] ET [29] Cosmic Explorer [30]
Kilohertz	Acoustic detectors	IGEC [31] IGEC2 [32] MiniGRAIL [12] Schenberg [13]

next decade like the Einstein gravitational wave Telescope (ET) and the Cosmic Explorer. For these detectors, a laser beam is the source of the EM coupling field. For a summary of various GW detectors, see Table 1.1. The relative phases (or difference in travel times) of the reflected beams are compared with the help of a beam splitter that recombines the beams. Any relative change in length of the two arms due to a passing GW is observed as an intensity variation at the beam splitter output.

1.3.1 Michelson Interferometer

The Michelson interferometer is a well-known experimental apparatus in physics and was first used to show the non-existence of the luminiferous ether in 1887. It is an extremely accurate instrument for measuring changes in the travel time of light in its arms. The simplest conceptual scheme is shown in Fig. 1.2. It consists of a monochromatic light source (a laser beam), whose light is sent on a beam-splitter which separates the light, into a beam travelling in one arm and a beam traveling in a second, orthogonal arm. At the end of each arm, we have totally reflecting mirrors. After travelling once back and forth, the two beams recombine at the beam splitter, and a part of the resulting beam goes to a photodetector, that measures the final output intensity. Let ω_L be the frequency of the laser so $k_L = \omega_L/c$ and $\lambda = 2\pi/k_L$ are the wave number and the wavelength of the laser. A given spatial component of the electric field of the input laser light is

$$E = E_0 e^{-i\omega_L t + i\boldsymbol{k}_L \cdot \boldsymbol{x}} \tag{1.35}$$

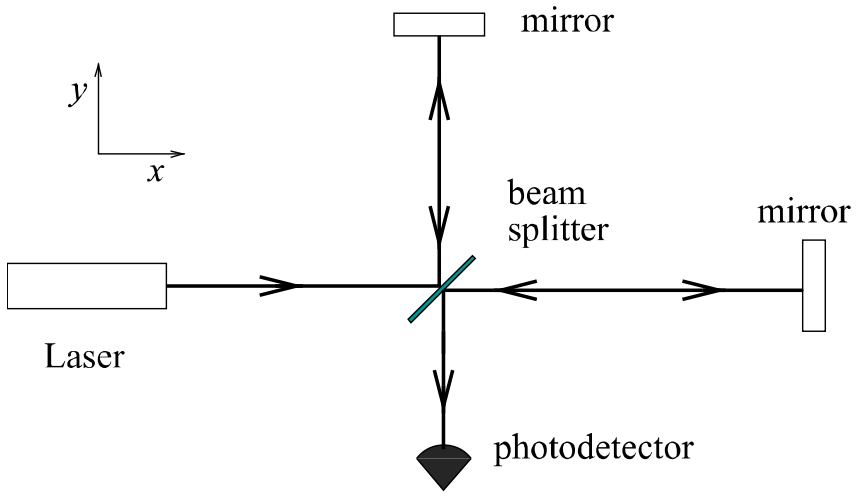


FIGURE 1.2: The layout of a simple Michelson interferometer.

Let's denote the length of the two arms by L_x and L_y. After the beam travels through the beam-splitter, the arms and then recombine again, when considering overall phase shift factors due to reflection and transmission, the power measured by the photodetector is proportional to

$$|E_{\text{out}}|^2 = E_0^2 \sin^2 \left(k_L(L_y - L_x)\right) \tag{1.36}$$

Thus, any variation in the length of the arms results in a corresponding variation of the power at the photodetector. The quantity inside the sine corresponds to an effective phase shift $\Delta\phi$ that can be broken down in two parts

$$\Delta\phi = \phi_0 + \Delta\phi_x(t) \tag{1.37}$$

where the phase ϕ_0 is a parameter controlled by the experimenter, chosen to have the best working point for the interferometer and $\Delta\phi_x(t)$ contains the effect of the GW. To see the full details of the formal derivation of $\Delta\phi_x$ in the TT gauge see [33]. It can be shown that the optimal length of the arms for a Michelson interferometer is $L = \lambda_{gw}/4$ or

$$L \equiv 750 \text{ km} \left(\frac{100 \text{ Hz}}{f_{gw}}\right) \tag{1.38}$$

Hence, to measure GWs with frequencies of order of ~ 100 Hz, the optimal choice would be an arm-length L of hundreds of kilometers. This is in practice very difficult. The alternative is therefore, to "fold" the optical path of light, making it bounce back and forth many times in each arm before recombining both beams. The solution adopted by modern interferometers like LIGO is transforming each arm into a *Fabry-Perot* (FP) cavity. Basically, photons are allowed to bounce back and forth hundreds of times between the mirrors increasing the time a GW can interact with the light resulting in better sensitivity.

1.3.2 Interaction with GWs in the proper detector frame

It is important now to understand the effects of a passing GW on our measurement device. For that, we need to go back to the mathematical description of GW and look into the detector's proper frame since this is the reference frame implicitly used by the experimenter. In this frame, the effect of the passing GW is a displacement of the test masses from their original position and, if the test masses are at a distance small compared to the reduced wavelength $\lambda\!\!\!^{-}_{GW}$ of the GW, the displacement is determined by Eq. (1.23) and the space-time metric can be taken as flat. From Eq. (1.19), a GW with a defined propagation direction $\hat{\boldsymbol{n}}$ can be expressed as

$$h_{ij}(t,\boldsymbol{x}) = \sum_{A=+,\times} e^A_{ij}(\hat{\boldsymbol{n}}) \int_{-\infty}^{\infty} df\, \tilde{h}_A(f) e^{-2\pi i f(t-\hat{\boldsymbol{n}}\cdot\boldsymbol{x}/c)} \tag{1.39}$$

For simplicity, we take $\boldsymbol{x} = \boldsymbol{0}$ as the localization of the detector. If the detector is sensitive only to GWs with a reduced wavelength much larger than its size ($2\pi f \hat{\boldsymbol{n}}\cdot\boldsymbol{x} = \hat{\boldsymbol{n}}\cdot\boldsymbol{x}/\lambda\!\!\!^{-} \ll 1$), the spatial dependence can be neglected, so that we can simply have

$$h_{ij}(t) = \sum_{A=+,\times} e^A_{ij}(\hat{\boldsymbol{n}}) \int_{-\infty}^{\infty} df\, \tilde{h}_A(f) e^{-2\pi i f t} = \sum_{A=+,\times} e^A_{ij}(\hat{\boldsymbol{n}}) h_A(t) \tag{1.40}$$

In general, it is useful to think of a GW detector as a linear system $h(t) = D^{ij}h_{ij}(t)$ so that the expected scalar calibrated output measured by the detector can be written as

$$h(t) = \sum_{A=+,\times} D^{ij} e^A_{ij}(\hat{\boldsymbol{n}}) h_A(t) \tag{1.41}$$

defining the *detector pattern function* $F_A(\hat{\boldsymbol{n}})$,

$$F_A(\hat{\boldsymbol{n}}) = D^{ij} e^A_{ij}(\hat{\boldsymbol{n}}). \tag{1.42}$$

Since the pattern function depends on the direction $\hat{\boldsymbol{n}}(\theta,\phi)$ and the detector's geometry, the measured strain can expressed in general as

$$h(t) = h_+(t)F_+(\theta,\phi) + h_\times(t)F_\times(\theta,\phi) \tag{1.43}$$

The response of an interferometer to a GW with an arbitrary direction is encoded in F_+ and $F_\times$. Considering the limit $\omega_{gw}L/c \ll 1$, the motion of the mirrors is governed by the equation of geodesics deviation (1.23). For two mirrors located at $\xi^j = (L,0,0),(0,L,0)$, the geodesic deviation reads

$$\ddot{\xi}_x = \frac{1}{2}\ddot{h}_{xx}L, \qquad \ddot{\xi}_y = \frac{1}{2}\ddot{h}_{yy}L. \tag{1.44}$$

These equations govern the change in the length of the x and y-arms of a Michelson interferometer, as well as the change in the length of a FP cavity. The relative shift between the x and y arms is therefore

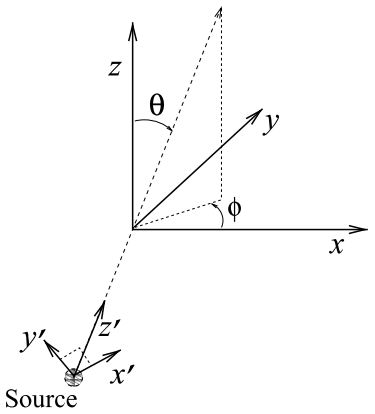


FIGURE 1.3: The geometry used in the computation of the pattern function computations. The arms of the interferometer are along the x and y axes.

driven by $\frac{1}{2}(\ddot{h}_{xx} - \ddot{h}_{yy})$. In other words, the detector tensor with arms along the $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}$ directions is

$$D_{ij} = \frac{1}{2}\left(\hat{\boldsymbol{x}}_i\hat{\boldsymbol{x}}_j - \hat{\boldsymbol{y}}_i\hat{\boldsymbol{y}}_j\right) \tag{1.45}$$

We have a frame (x, y, z) (proper detector frame) such that the arms of the interferometer are along the x and y axes. Then, we introduce a second frame of reference, the frame of the source (x', y', z'), such that the propagation direction of the GW coincide with the z' axis. With respect to the (x, y, z) frame, the z' axis has polar angles θ and ϕ, as shown in Fig. 1.3. The polarizations h_+ and $h_\times$ are defined with respect to (x', y'), so in the (x', y', z') frame the GW has the form

$$h^{'}_{ij} = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}_{ij} \tag{1.46}$$

The rotation transformation that brings the (x', y', z') frame onto the (x, y, z) frame is given by a rotation by an angle θ around the y axis followed by a rotation by an angle ψ around the z axis

$$\mathcal{R} = \begin{pmatrix} \cos(\phi) & \sin(\phi) & 0 \\ -\sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & \cos 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix} \tag{1.47}$$

The GW in the (x, y, z) frame is then given by the tensor transformation law $h_{ij} = \mathcal{R}_{ik}\mathcal{R}_{jl}h^{'}_{kl}$. Upon calculation we obtain

$$h_{xx} = h_+\left(\cos^2\theta\cos^2\phi - \sin^2\phi\right) + 2h_\times\cos\theta\sin\phi\cos\phi \tag{1.48}$$

$$h_{yy} = h_+\left(\cos^2\theta\sin^2\phi - \cos^2\phi\right) - 2h_\times\cos\theta\sin\phi\cos\phi \tag{1.49}$$

resulting in

$$\frac{1}{2}\left(h_{xx} - h_{yy}\right) = \frac{1}{2}\left(1 + \cos^2\theta\right)\cos 2\phi + h_\times\cos\theta\sin 2\phi \tag{1.50}$$

and therefore

$$F_+(\theta,\phi) = \frac{1}{2}\left(1+\cos^2\theta\right)\cos 2\phi \tag{1.51}$$

$$F_\times(\theta,\phi) = \cos\theta \sin 2\phi \tag{1.52}$$

These functions represent the sensitivity of the Michelson interferometer. One peculiar feature is that GW interferometers have blind directions. For instance, for a GW with a plus polarization the direction $\phi = \pi/4$ is blind.

1.3.3 Real GW interferometers

In the real experiment, there are many idealizations that do not hold and many other technical issues that have to be accounted for. First, any laser beam has a profile in the transverse direction since it is subject to diffraction. If at some point in space, a photon of wavelength λ_L and longitudinal momentum $p = \hbar/\lambda\!\!\!^{-}_L$ is localized within a transverse width $\Delta x_\perp = a$, by the Heisenberg principle, it has an uncertainty on the transverse momentum $\Delta p_\perp \sim \hbar/a$ so the beam will widen, filling a cone of angle $\Delta\theta = \Delta p_\perp/p \sim \lambda\!\!\!^{-}_L/a$. After traveling a longitudinal distance x the beam will become larger, in the transverse direction by $x\Delta\theta \sim x\lambda\!\!\!^{-}_L/a$. For interferometers like LIGO and VIRGO, the wavelength of the laser is $\lambda_L \sim 1\,\mu$m. For $x = 4$ km and $\lambda_L = 1\,\mu$m , we have $a \equiv 2.5$ cm, the broadening of the light becomes important just in a single one-way trip. Furthermore, for FP cavities where the beam is supposed to perform hundreds of round trips, the beam would widen even further. When considering the Gaussian nature of laser beams, it can be shown [34] that the surface of constant phase is the portion of a spherical surface, hence, the mirrors inside the cavities should be spherical surfaces.

Second, the experimenter has to choose a point of operation, i.e. choose an initial phase ϕ_0 from which the phase due to any gravitational perturbation will be measured. In Michelson interferometers, the GW produces a phase shift of the light $\Delta\phi_{gw}$. To extract this phase information from the output of the detector, as shown in Fig. 1.4, a ϕ_0 has to be chosen so that any variation in power $P = P_0 \sin^2\Delta\phi$ around that point can be accurately measured. At first, one might think that choosing $\phi_0 = \pi/4$ yields the best working point since the derivative $\partial P/\partial\phi_0$ is maximum and the sensitivity to small displacements $\phi_0 \to \phi_0 + \Delta\phi_{gw}(t)$ is highest, However, at this point the measurement is very sensitive to fluctuations in laser power P_0. The usual practical solution to this problem is to choose the *dark fringe* point, i.e. when the signal is absent, the instrument records zero output. This make the instruments insensitive to calibration uncertainties. Unfortunately, at this point $P = 0$ and $\partial P/\partial\phi = 0$ and because the change in power due to GW is $\Delta P \sim \mathcal{O}(h^2)$, an effect quadratic in h would effectively be invisible. The standard solution to this problem is *phase modulation* of the input laser light using a Pockels cells. In a nutshell, a Pockel cell is a crystals or block of dielectric material whose index of refraction depends on an applied electric field $E_{\rm ap} = |E_{\rm ap}|\cos\Omega_{\rm mod}t$ whose overall effect is to break down the effect of the GW of frequency f_{gw} into a carrier wave and side-bands that encode the signal in a term that oscillates as $\cos(\Omega_{\rm mod}t - \alpha)$ that no longer competes with the fluctuation of the laser at a frequency f_{gw} but rather

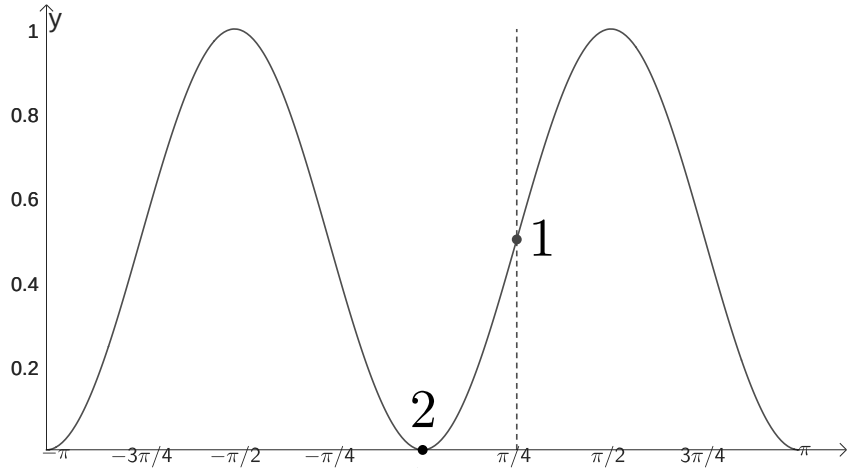


FIGURE 1.4: The power $P(\phi)/P_0$. The naive working point is marked as 1, and the dark fringe as 2.

with the fluctuations of the laser at a frequency f_{mod} which usually have a $1/f$ dependency, therefore making it very small.

One further improvement is the so-called *power recycling*. Basically, since the dark fringe point was chosen for the carrier wave, in the absence of GWs no light emerges from the beam-splitter in the direction of the photodetector at the carrier frequency. This means that all the light at frequency ω_L that circulates in the arms is eventually reflected by the beam-splitter back toward the laser and essentially wasted. To counteract shot noise (Sec. 1.3.4.1), it is desirable to have the highest possible power laser intensity circulating in the arms. To increase it, the idea is to "recycle" the light that comes back toward the laser, placing a mirror that reflect the light back toward the beam-splitter effectively creating a new FP cavity . If the cavity is arranged so that it is resonant for the input laser light, the total intensity of the light that circulates in the interferometer is enhanced, even by a factor $\sim \mathcal{O}(100)$. Finally, we also need an *output power cleaner*. Even if the initial beam has been accurately prepared to very well defined low modes thank to the input mode cleaner, imperfections in the mirrors or misalignments can regenerate higher modes inside the interferometer. These higher modes are not in the dark fringe and therefore produce undesirable noise that lower the contrast of the output. The output mode cleaner, placed between the beam-splitter and the photodetectors filters out these higher modes. Putting it all together, we arrive at an optical layout like the one in the illustration shown in Fig. 1.5.

1.3.4 Noise sources

It is impossible to obtain a clear reading from a measurement instrument. There are always limitations related to the chosen experimental methods or as we see in the next sections, to the physical limits of accuracy imposed by physical effects inherent to nature.

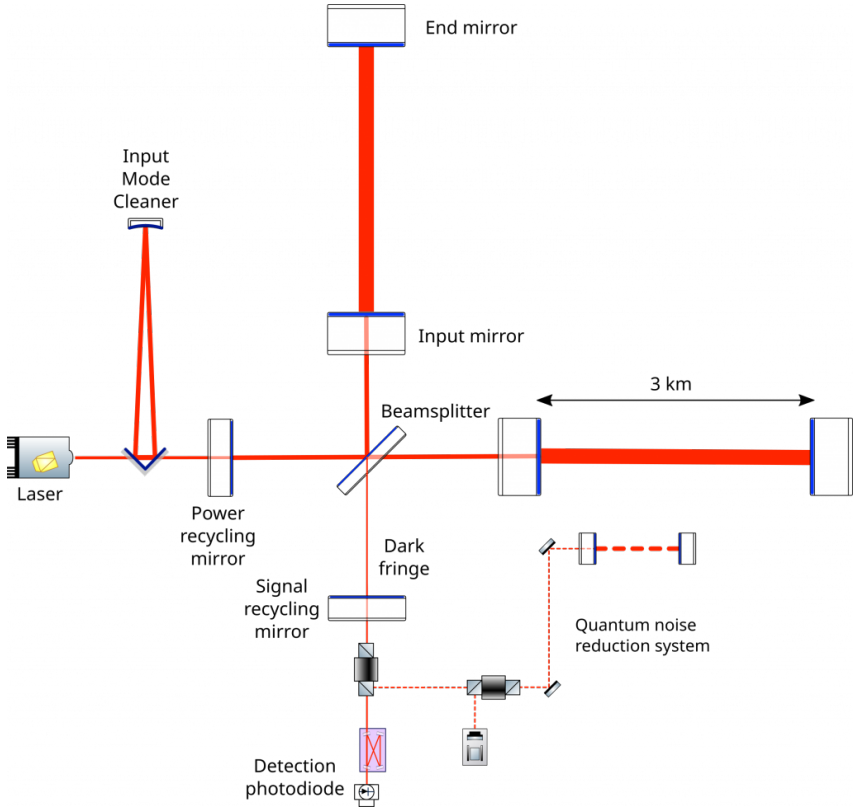


FIGURE 1.5: The basic layout of a GW interferometer with dual recycling and Quantum Noise reduction.

1.3.4.1 Shot noise

This noise is originated from the fact that the laser light comes in discrete quanta. Let N_γ be the number of photons that arrives on the photodetector in an observation time T, the average power measured at the photodetector during the observation time is

$$P = \frac{1}{T} N_\gamma \hbar \omega_L \tag{1.53}$$

In reality, measuring the average power is the same as counting the number of photons that arrived in a time T. Such a phenomenon follows a Poisson distribution which in the limit of large N becomes Gaussian with a standard deviation $\sqrt{N}$. Working at a generic point ϕ_0 and neglecting modulation of laser light, this produces a fluctuation in the observed power given by

$$\Delta P_{\rm shot} = \left(\frac{\hbar \omega_L}{T} P \right)^{1/2} = \left(\frac{\hbar \omega_L}{T} P_0 \right)^{1/2} |\sin \Delta\phi_0| \tag{1.54}$$

On the other hand, the variation of power due to a GW is

$$\Delta P_{\rm gw} = \frac{P_0}{2} |\sin 2\phi_0| \Delta\phi_{\rm Mich} \tag{1.55}$$

For simplicity, we consider a periodic GW with frequency f with only the plus polarization and coming from an optimal orientation, and taking $2\pi f L/c \ll 1$, the amplitude of the phase shift $\Delta\phi_{\rm Mich}$ is given

by

$$|\Delta\phi_{\rm Mich}| = \frac{4\pi L}{\lambda_L} h_0, \tag{1.56}$$

The signal-to-noise ratio for this case is

$$\frac{S}{N} = \frac{\Delta P_{\rm gw}}{\Delta P_{\rm shot}} = \left(\frac{P_0 T}{2\hbar\omega_L}\right)^{1/2} \frac{4\pi L}{\lambda_L} h_0 |\cos\phi_0| \tag{1.57}$$

It is often convenient to express this intensity as the strain sensitivity $S_n^{1/2}(f)$ as

$$S_n^{1/2}(f) = \frac{T^{1/2}}{(S/N)} h_0 = \frac{\lambda_L}{4\pi L |\cos\phi_0|} \left(\frac{\hbar\omega_L}{P_0}\right)^{1/2} \tag{1.58}$$

which is inversely proportional to the laser power P. Then, power recycling indeed decreases the level of shot noise.

1.3.4.2 Radiation pressure

When photons reflect back from the mirrors, they exert a pressure on the mirror itself caused by the change of momentum from $+p$ to $-p$, so it transfers a total momentum $2p$. Since the photon energy is $E_\gamma = p/c$, the force that a beam of power P exerts on the mirror is $F = 2P/c$. The RMS fluctuation of the force in a time T is related to the power fluctuations by

$$\Delta F = \frac{2\Delta P}{c} = 2\left(\frac{\hbar\omega_L P_0}{c^2 T}\right)^{1/2} \tag{1.59}$$

So, as we can see, increasing the power of the laser increases the radiation pressure. Then, it is important to find the balance between lowering shot noise and further increasing radiation pressure. To assess this, the mirrors have suspension mechanisms that compensate for the motion caused by the incident photons. However, they introduce further sources of noise.

The combined effect of shot noise and radiation pressure is known as *optical read-out* noise. The shot noise contribution is proportional to $P_0^{-1/2}$ while the radiation pressure to $P_0^{1/2}$. The photons impart non-deterministically a recoil to the mirror in the form of radiation pressure and this recoil disturbs the measure that we perform. The optimal strain sensitivity corresponding to both noises defines the *standard quantum-limit* (SQL)

$$S_{\rm SQL}^{1/2}(f) = \frac{1}{2\pi f L}\sqrt{\frac{8\hbar}{M}} \tag{1.60}$$

It is a macroscopic manifestation of the uncertainty principle. However, the uncertainty does not impose a limit on the accuracy of position, but only on the accuracy of simultaneous measurement of conjugate variables and it is possible to go beyond the standard quantum limit using quantum non-demolition techniques which are important for second generation advanced interferometers [35]. These techniques include squeezed state injection, ponderomotive squeezing and variational readout, and optical springs.

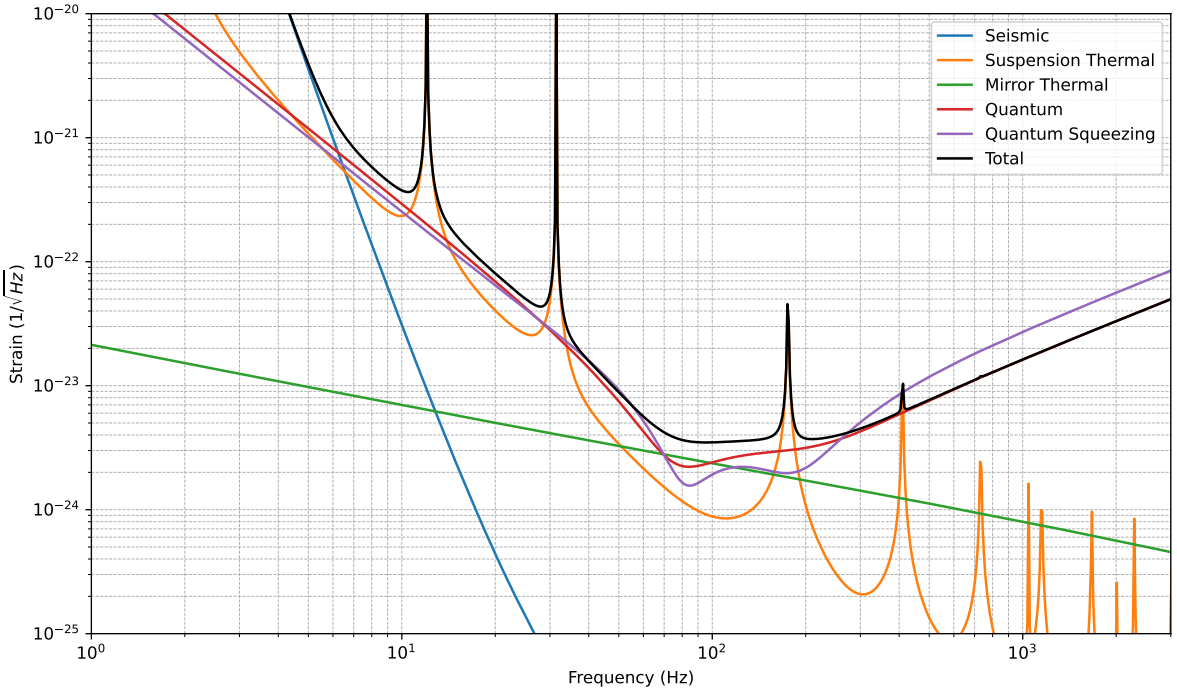


FIGURE 1.6: Design sensitivity of KAGRA [36].

1.3.4.3 Displacement Noise

There is motion of the test masses that have nothing to do with GWs or optical systems. They depend on many technical issues such as properties of materials, details of the suspension mechanisms, etc. The main source of such noises is Earth's ground which is in continuous motion with amplitudes of order of microns, the so called *seismic noise*. In the region of $\sim$ 10 Hz, human activity such as local traffic, trains, etc, creates another source of vibration.

Furthermore, *Newtonian noise* or "gravity gradient noise" originates in the Newtonian gravitational forces of objects that are moving which causes a gravitational force that changes in time. The most important component is induced by micro-seismic noise, which produce mass density fluctuations of the gravitational field of the Earth. In principle, detectors can be isolated to attenuate seismic noise but Newtonian noise will always provide a great limitation for the low frequency range of detection.

1.3.4.4 Thermal Noise

It arises from random energy transfer between the detector and the environment, inducing vibrations in both the suspension system and the mirror. In the *suspension thermal noise*, we have pendulum thermal fluctuations that induce a swinging motion in the suspensions and a horizontal displacement of the mirrors, vertical thermal fluctuations that add a vertical motion of the suspensions, violin modes that are described in terms of fluctuations of the normal modes of the wires and are responsible for the spikes pass 300 Hz. We also have *test-mass thermal noise* which is the fluctuation between the test masses themselves. Brownian motion of the mirrors in their atoms, thermo-elastic fluctuations

originated through the expansion of the material and thermo-refractive fluctuations that come from the temperature dependency of the refraction index of the coatings. The design sensitivity plot of the Kamioka Gravitational Wave Detector (KAGRA) and the spectral noise density of various noise sources is shown in Fig. 1.6.

1.4 Data Analysis

Now that we are aware of the possible source of noises in detectors, we can introduce the optimum filtering techniques that must be applied to the detector output. The importance of this procedure stems from the fact that, with existing detectors and with reasonable estimates of the GW signal, we expect that the GW signal will be buried into a much larger noise. This is a typical problem in many fields, e.g. radio astronomy for application to pulsar searches, and standard filtering techniques have been developed. We will see how these techniques are adapted to the problem of GW detection.

1.4.1 Matched Filtering

The output of any real detector $s(t)$ can be modelled as a linear system composed by a strain component $h(t)$ and noise $n(t)$ as

$$s(t) = h(t) + n(t). \tag{1.61}$$

How can the signal be dug out the GW signal from a much larger noise? As exemplified by radio astronomy, it can be done by knowing, at least to some level of accuracy, the form of $h(t)$. To see this, we can take the previous equation, multiply it $h(t)$, integrate over an observation time T and average over it

$$\frac{1}{T}\int_0^T dt\, s(t)h(t) = \underbrace{\frac{1}{T}\int_0^T dt\, h^2(t)}_{I_1} + \underbrace{\frac{1}{T}\int_0^T dt\, n(t)h(t)}_{I_2}, \tag{1.62}$$

it is clear that I_1 is positive definite and its value averaged over a time T is of order of T rendering $I_1 \sim h_0^2$, where h_0 is the characteristic amplitude for the oscillating function $h(t)$. In contrast, the factor in the integrand of I_2 are uncorrelated. The quantity $n(t)h(t)$ is oscillating, and its integral will grow as $T^{1/2}$ for large T, so $I_2 \sim (\tau_0/T)^{1/2} n_0 h_0$, where n_0 is the characteristic amplitude of the oscillating function $n(t)$ and τ_0 is a typical characteristic time. Thus, in the limit $T \to \infty$, I_2 averages to zero. That is the basic idea of "filtering" which mathematically, can be optimised in order to obtain the highest possible value of the signal-to-noise (SNR) ratio. For that, we define

$$\hat{s} = \int_{-\infty}^{\infty} dt\, s(t)K(t), \tag{1.63}$$

where $K(t)$ is called of the *filter* function. In the following, we assume that we know the shape of $h(t)$. Since the filter function is going to be taken to "match" the signal we look for, the technique is called "matched filtering". The SNR is defined as S/N, where S is the expected value of $\hat{s}$ when the signal is

present, and N is the Root Mean Square (RMS) value of $\hat{s}$ when the signal is absent. Since it is assumed that the noise component is Gaussian, $\langle n(t)\rangle = 0$, we have

$$S = \int_{-\infty}^{\infty} dt\, \langle s(t)\rangle K(t) = \int_{-\infty}^{\infty} dt\, h(t)K(t) \tag{1.64}$$

which in the frequency space is

$$S = \int_{-\infty}^{\infty} df\, \tilde{h}(f)\tilde{K}^*(f) \tag{1.65}$$

and

$$\begin{aligned} N^2 &= \langle \hat{s}^2(t)\rangle_{h=0} \\ &= \int_{-\infty}^{\infty} dtdt'\, K(t)K(t')\langle n(t)n(t')\rangle \\ &= \int_{-\infty}^{\infty} dtdt' K(t)K(t') \int_{-\infty}^{\infty} df df' e^{2\pi i f t - 2\pi i f' t'} \langle \tilde{n}^*(f)\tilde{n}(f')\rangle \\ &= \int_{-\infty}^{\infty} df\, \frac{1}{2} S_n(f)|\tilde{K}(f)|^2 \end{aligned} \tag{1.66}$$

and therefore

$$\frac{S}{N} = \frac{\int_{-\infty}^{\infty} df \tilde{h}(f)\tilde{K}^*(f)}{\left[\int_{-\infty}^{\infty} df (1/2) S_n(f)|\tilde{K}(f)|^2\right]^{1/2}} \tag{1.67}$$

To find the filter that maximizes the SNR, it is customary to define the scalar product between to real functions $A(t)$ and $B(t)$ as

$$(A|B) = \Re \int_{-\infty}^{\infty} df \frac{\tilde{A}^*(f)\tilde{B}(f)}{(1/2)S_n(f)} = 4\Re \int_0^{\infty} \frac{\tilde{A}^*(f)\tilde{B}(f)}{S_n(f)} \tag{1.68}$$

where Re denotes the real part and $A(t)$ and $B(t)$ are taken to be real functions. Then, the SNR can be written as

$$\frac{S}{N} = \frac{(u|h)}{(u|u)^{1/2}} \tag{1.69}$$

where $u(t)$ is the function whose Fourier transform is

$$\tilde{u}(f) = \frac{1}{2} S_n(f)\tilde{K}(f) \tag{1.70}$$

In this form, the optimization problem can be interpreted as searching for the vector of unit norm $\hat{n}u/(u|u)^{1/2}$, such that its scalar product with the vector h is maximum. This can be achieved by choosing $\hat{n}$ and h to be "parallel", in other words

$$\tilde{K}(f) = \text{const}\frac{\tilde{h}(f)}{S_n(f)}. \tag{1.71}$$

Plugging this result into Eq. (1.70) we get $\tilde{u} = \text{const} \times \tilde{h}$. Inserting this into Eq. (1.69) we get the optimal SNR value

$$\left(\frac{S}{N}\right)^2 = 4\int_0^{\infty} df \frac{|\tilde{h}(f)|^2}{S_n(f)}, \tag{1.72}$$

which is completely general and independent of the form of $\tilde{h}(f)$.

1.5 Astrophysical sources

For purposes of detection, one can usefully classify sources in four broad categories [37]: 1) short-lived and well defined, for which coalescence of a compact binary system is the canonical example; 2) short-lived and a priori poorly known, for which a supernova explosion is the canonical example; 3) long-lived and well defined, e.g., continuous waves from spinning neutron stars; and 4) long-lived and stochastic, e.g., primordial gravitational waves from the Big Bang. For existing and upcoming terrestrial detectors, the most promising category is the first. Detectable event rates for compact binary coalescence can be estimated with the greatest confidence and imply highly likely discovery by Advanced LIGO and Virgo detectors.

For future spaced-based detectors, which can probe to lower frequencies, the pre-coalescence phase of galactic binaries NS-NS is accessible, at the same time that coalescence of binary super-massive black holes (SMBHs), e.g., from galaxy mergers should be detectable [38]. Similarly, pulsar timing arrays can potentially detect a stochastic astrophysical background from the superposition of cosmologically distant SMBH binary systems [39] at still lower frequencies ($\sim$ nHz). For simplicity, we show here four main types of gravitational radiation sources: compact binary coalescences, bursts, continuous waves and the stochastic background.

1.5.1 Compact Binary Coalescences (CBCs)

Binary star systems are common in our galaxy, but only a tiny fraction evolve into compact binary systems that can merge within a Hubble time [40]. For a CBC to occur, both stars must be massive enough to collapse into compact objects without destroying each other, shedding too much mass, or undergoing disruptive perturbations.

CBCs are the most promising sources for GW detection, particularly for advanced detectors like LIGO and VIRGO, which currently identify hundreds of CBC events annually. The coalescence of two neutron stars (NS-NS), a neutron star and a black hole (NS-BH), or two black holes (BH-BH) produces strong GWs that current detectors can observe. The comparison of the predicted waveforms [41] with observed ones constitutes the strongest test of GR. This is especially significant when multiple detectors are involved, enabling the separation of waveform polarizations. To exploit this opportunity, we must be able to follow closely the signal with a template. This means, that for a given value of the parameters of the binary system (time of coalescence, masses, spin, etc) one must know the waveform accurately. Fig. 1.7 illustrates the event GW150914 as detected by the LIGO Hanford and Livingston detectors. The comparison between the observed waveform and numerical relativity reconstructions are remarkable and demonstrates the high degree of accuracy, as indicated by the minimal residuals.

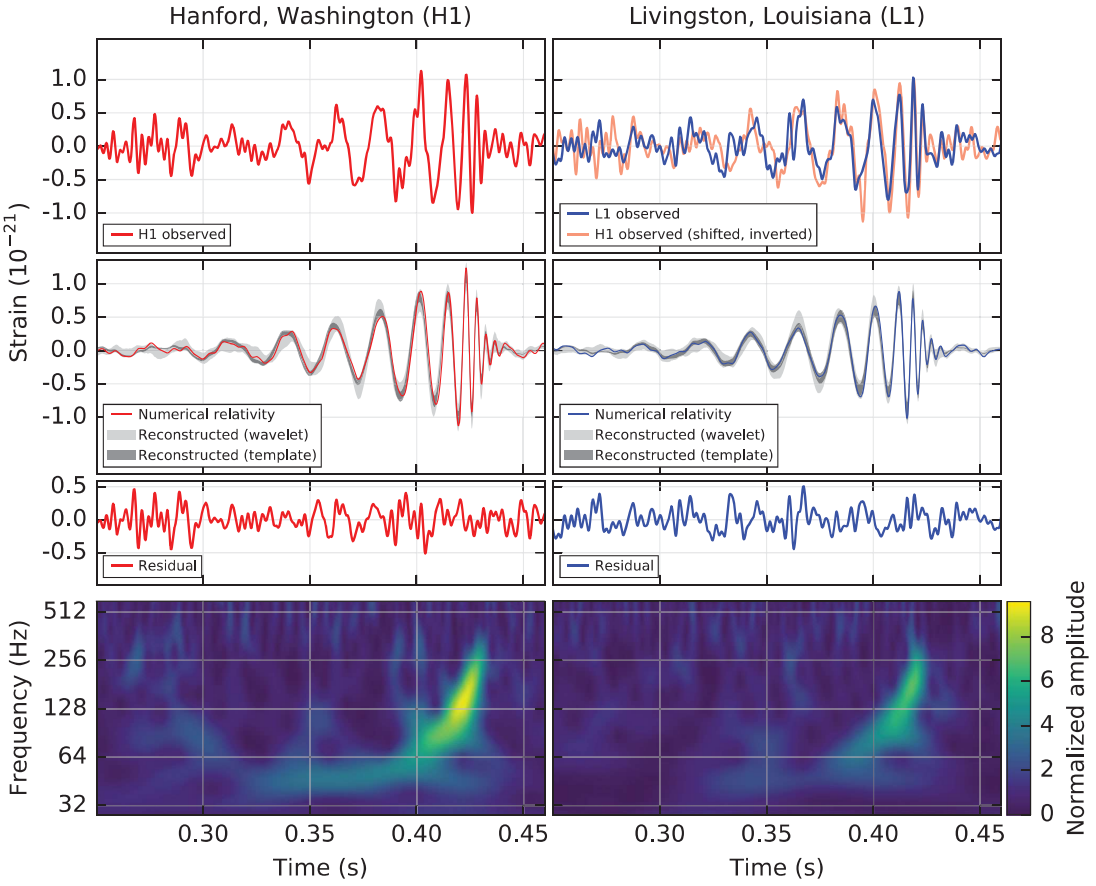


FIGURE 1.7: GW event GW150914 detected by the LIGO Hanford (H1) and Livingston (L1) detectors [3]. Top Row: Observed strain data from H1 (left) and L1 (right), with L1 data shifted and inverted for comparison. Second Row: Comparison of the observed waveform with numerical relativity, wavelet, and template reconstructions for both detectors. Third Row: Residuals after subtracting the reconstructed signal from the observed data for both detectors. Bottom Row: Time-frequency representation (Q-transform) of the strain data, showing the signal's frequency increasing over time. The numerical relativity waveform models a binary black hole merger, with shaded areas indicating 90% confidence regions. The high overlap between the numerical relativity results and observed waveforms confirms the accuracy of the reconstruction methods.

The merger of two compact massive objects, such as NS or BH, into a single BH, can be divided into three distinct stages: inspiral, merger, and ringdown. During the inspiral stage, gravitational waveforms can be accurately described using analytic expressions known as perturbative post-Newtonian (PN) approximations. The merger stage, characterized by strong relativistic effects, requires fully non-linear numerical relativity calculations. Finally, the ringdown stage of the newly formed black hole is expected to be relatively simple again.

Thus the early phases of the inspiral stage [42] should provide a well understood post-Newtonian system, from which stellar masses and spins can be determined. With these parameters determined, one can then make detailed comparisons of observations in the merger stage with numerical predictions for those parameters. The ringdown mode frequencies and damping times are primarily governed by the total mass and spin of the final black hole, allowing clean and analytic comparisons to the inspiral stage, largely independent of the merger uncertainties.

Coalescences involving neutron stars offer the potential for probing the neutron star equation of state via distortions of the detected waveform away from that expected for two point masses, because of tidal

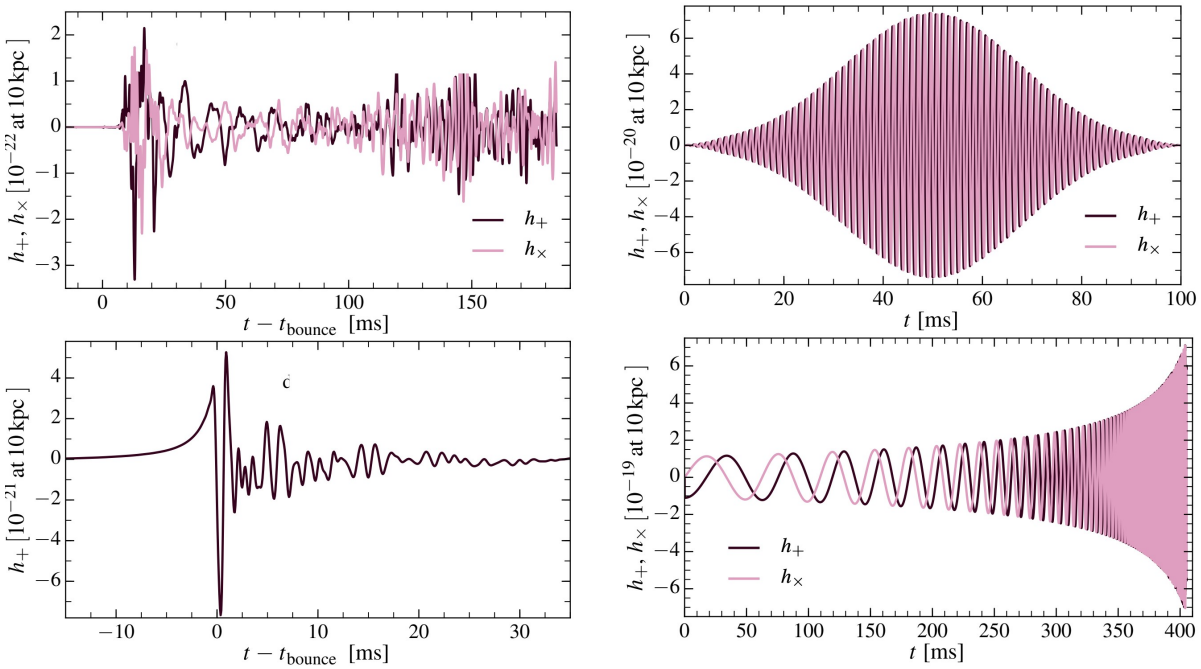


FIGURE 1.8: Inventory of waveforms from CCSNe adapted from [43] Left panel shows the strain for a neutrino (top) and rotation (bottom) driven core collapse a seen from an equatorial observer at a distance of 10 kpc. Right panels show the strain from a bar-mode (top) and disc fragmentation instability (bottom), as seen by a polar observer at 10 kpc, calculated to leading (quadrupolar) order from analytical modellisation of the instabilities.

disruption of one or both stars. The effects are expected to be small, however, and their detection dependent on the detector sensitivity at the highest frequencies in the detector bands.

1.5.2 Bursts

GW bursts customarily refer to transients of poorly known or unknown phase evolution. Although the algorithms used to search for bursts are sensitive to high- SNR, well predicted waveforms such as from NS-NS coalescence, they are necessarily less sensitive than matched-filter approaches, where known phase evolution can be exploited. More generic transient algorithms must be used for burst sources, such as supernovae, because of uncertain dynamics in these violent processes and because of uncertain but almost certainly varying initial stellar conditions.

It is important to note that a spherically symmetric explosion (or implosion) does not lead to GWs in GR (no monopole term). To be detected via GWs then, a supernova presumably needs to exhibit some asymmetry. The fact that many pulsars formed in supernovae have large measured speeds relative to their neighbors [44] strongly suggests that some supernovae do exhibit substantial non-spherical motion, perhaps as a result of dynamical instabilities in rapidly rotating, massive progenitor stars. One recently appreciated mechanism for potentially strong gravitational wave emission during core-collapse supernovae is hydrodynamical oscillation of the protoneutron star core [45]. With GW detection now on the horizon, much work has gone into detailed theoretical waveform modeling, as depicted in Fig. 1.8, and simulations of the supernova process to predict possible resulting waveforms. As one might imagine, this

violent process, in which strong magneto-hydrodynamics, nuclear physics and GR are all important, is a formidable challenge to simulate. Indeed, it has proven challenging to reproduce in these simulations the spectacular explosions that we associate with supernovae [38]. Nonetheless, this recent work has given new insights into the strength and spectral content one might expect from supernovae. Unfortunately, predictions of strength remain subject to large uncertainties. For scale, consider a supernova a distance r away in our galaxy that emits energy E in GWs, with a characteristic duration T and characteristic frequency f. One expects [38] a detectable strain amplitude at the Earth of about

$$h \sim 6 \times 10^{-21} \left(\frac{E}{10^{-7} M_\odot c^2}\right)^{1/2} \left(\frac{1 \text{ ms}}{T}\right) \left(\frac{1 \text{ kHz}}{f}\right) \left(\frac{10 \text{ kpc}}{r}\right) \tag{1.73}$$

For the nominal (but not necessarily accurate) values of r, E, T and f in this expression, the initial LIGO and Virgo interferometers should have been able to detect a galactic supernova in GWs. But no supernova was detected electromagnetically in our galaxy during initial LIGO and Virgo data taking, which is unsurprising, giving their expected low occurrence rate. With the expected order of magnitude improvement in sensitivity of the advanced detectors, a galactic supernova with 100 times smaller energy or a supernova with the same energy ten times further away would be accessible. Note, however, that until one reaches the Andromeda galaxy ($\sim$ 780 kpc), there is relatively little additional stellar mass beyond the edge of the Milky Way, with nearby dwarf galaxies contributing only a few percent additional mass. (Nonetheless, the most recent known nearby supernova SN1987A was in the Large Magellanic Cloud at $\sim$ 50 kpc.) One intriguing scenario in which a core collapse supernova could be seen in GWs to much larger distances is via a bar mode instability [38], in which differential rotation in a collapsing star leads to a large, rapidly spinning quadrupole moment, generating waves detectable from well outside our own galaxy [46]. Another type of instability (r-mode) may develop in the birth of a neutron star, but its lifetime is expected to be long enough, that it will be discussed below in the category of continuous wave sources.

Another potential transient source of poorly known gravitational waveform shape is the sudden release of energy from a highly magnetized neutron star (magnetar). Although "ordinary" neutron stars are characterized by extremely strong surface magnetic fields ($\sim 10^{10}$ G), many magnetars appear to have fields $\sim 100 - 1000$ times still stronger, implying enormous pent-up magnetic energy. It is thought that soft gamma ray repeaters (SGRs) and anomalous X-ray pulsars (AXPs) are different observational manifestations of the same underlying system a highly magnetized star which sporadically converts magnetic field energy into radiation [47]. Whether this process involves rupture of the neutron star crust, vortex rearrangement in a core superconducting fluid, or some other process, is not yet well understood. Especially dramatic instances are superflares, such as the December 2004 flare from SGR 1806-20, in which $\sim 10^{39}$ J of electromagnetic energy was released [48]. This radiation release from $\sim$ 10 kpc away disturbed the Earth's ionosphere sufficiently to disrupt some radio communications [49]. How much gravitational wave energy might be released in such events is unclear, although it has been proposed that the energy released into neutron star crustal vibrations could be comparable to that released electromagnetically [49], in which case gravitational radiation due to those vibrations could be substantial. For scale, the magnetic energy stored in a neutron star with surface field of 10^{15} G

is $\mathcal{O}(10^{40}J)$, assuming an internal field no larger than the surface field. If the star had still stronger internal fields, the energy would be still larger. Given the uncertainties in the mechanism leading to these enormous radiation releases, it is hard to be confident of predicted waveforms. Hence generic GW transient algorithms are appropriate in searching for flares. Another possible transient source is emission of bursts of gravitational radiation from "cosmic string cusps" [50]. Cosmic strings might be defects remaining from the electroweak (or earlier) phase transition or possibly primordial superstrings redshifted to enormous distances. In either model, kinks in these strings would travel at the speed of light with an isotropic distribution of directions, generating a model-dependent gravitational wave spectrum that is collimated along the direction of cusp travel. According to this idea, one would expect a cosmological background of GW bursts, that might be detected individually. As discussed below, this same model could lead to a steady-state, lower-level stochastic background from cusp radiation from more distant reaches of the universe. A general consideration in burst searches is the energy release implicit for a given source distance and detectable strain amplitude. As the distance of the source increases, the energy required for its waves to be detectable on Earth increases as the square of the distance. Specifically, rewriting Eq. (1.73), one obtains the relation

$$E \sim 3 \times 10^{-3} M_{\odot} c^2 \left(\frac{h}{10^{-21}}\right)^2 \left(\frac{T}{1\text{ ms}}\right) \left(\frac{f}{1\text{ KHz}}\right) \left(\frac{r}{10\text{ Mpc}}\right)^2 \tag{1.74}$$

Hence for a source distance much beyond 10 Mpc and for initial LIGO/Virgo sensitivities to transients, one needs sources emitting significant fractions of a solar mass in gravitational radiation in frequency bands accessible to terrestrial detectors, such as is expected in the case of coalescing binary systems.

1.5.3 Continuous Gravitational Waves (CWs)

Continuous Gravitational waves sources emit radiation over very long observational periods and are expected to have well modelled quasi-monochromatic signals . In the frequency range of ground-based interferometric detectors, rapidly rotating neutron stars are the most likely sources of periodic, persistent gravitational waves [51]. The inner cores of massive stars can have high angular momentum and theory suggests that neutron stars could be born rotating at near the maximum value that they can endure without breaking up, which is of order of $\sim 10^3$ Hz [52]. This is much faster than the observed spin rates of the majority of known pulsars. One explanation for this discrepancy is that angular momentum is removed through gravitational radiation. Rotating neutron stars could radiate gravitationally due to asymmetries. Deformations could be introduced due to build-up of strain in the crust, through the magnetic field or from accretion. Approximations of the characteristic amplitude h_c of GW from rotating neutron stars take the form

$$h_c \sim \left(\frac{G}{c^4}\right)\left(\frac{If^2\delta}{d}\right), \tag{1.75}$$

where I is the neutron star's moment of inertia and f is the GW frequency. The distance to the source is given by d and δ measures the deviation of the star from axis-symmetry. The last equation shows that the strength of radiation depends on how rapidly the source rotates and the degree of distortion. Although the amplitudes are expected to be weak $h_c \sim 10^{-26}$, matched-filtering can be applied over long observational

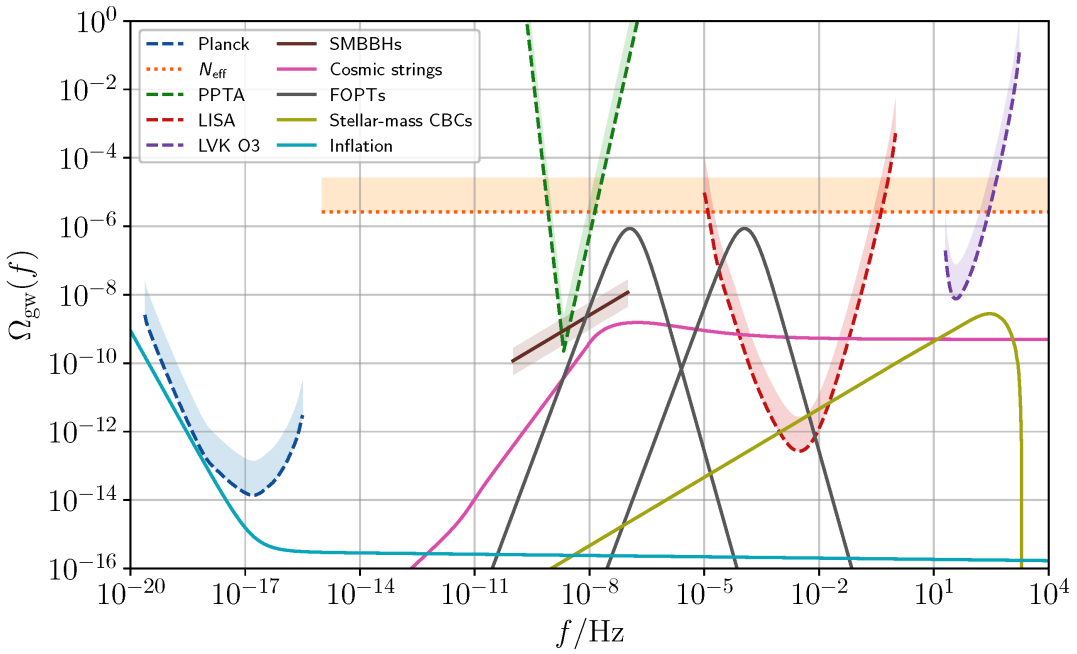


FIGURE 1.9: An overview of potential GW background signals across the frequency spectrum taken from [54]. The light blue curve represents the prediction for single-field slow-roll inflation with a canonical kinetic term. The pink curve illustrates a GW background from Nambu-Goto cosmic strings. The brown curve depicts a GW background from inspiraling supermassive binary black holes (BBHs), with the amplitude and shaded region corresponding to the common noise process in the NANOGrav 12.5-year data set [14]. The two gray curves represent GW backgrounds generated by first-order phase transitions at the electroweak scale ($\sim$ 100 GeV) and the Quantum Chromo-Dynamics (QCD) scale ($\sim$ 200 MeV), respectively. The yellow curve shows a GW background generated by stellar-mass compact binaries, based on the mass distributions and local merger rates inferred from LVK detections. The dashed curves indicate various observational constraints. The dotted curve represents the integrated constraint from measurements of $N_{\rm eff}$, which cannot be directly compared with the frequency-dependent constraint curves but is included here for illustrative purposes.

period to build up the signal-to-noise-ratio. This technique is complicated by Earth's rotational and orbital motion, as well as pulsar spin-downs. Proposed mechanisms for producing distortions include the Chandrasekhar-Friedman-Schutz instability. This is driven by GW back reaction that creates and maintains hydrodynamic waves in neutron star's fluid components, known as *r-modes*, which propagate in the opposite direction to that of the star's rotation, producing gravitational radiation. We are going to delve into a lot more detail on this in in chapter 2.

1.5.4 The Stochastic Background

Stochastic GWs arise from a superposition of incoherent sources. While a cosmological background from primordial gravitational waves created in the Big Bang are a natural possible source [53], other isotropic possibilities are from cosmic strings and from very distant mergers of neutron stars or of supermassive black holes (accessible to space-based detectors). Non-isotropic sources in the band of terrestrial detectors could include the superposition of pulsar radiation from, say, the Virgo Cluster. Over very long time scales, gravitational radiation from an accreting neutron star could also appear stochastic, as the phase of the narrowband signal wanders. A primordial isotropic gravitational wave background is predicted by most cosmological theories, although the predicted strengths of the background vary enormously. It

is customary [46, 55] to parametrize the background strength vs. frequency f by its energy density per unit logarithm normalized to the present-day critical energy density $\rho_{\rm crit} = 3H_0^2c^2/8\pi G$ of the universe, where H_0 is Hubble's constant, taken here to be 70.5 km/s/Mpc [56]

$$\Omega_{\rm gw}(f) = \frac{1}{\rho_{\rm crit}} \frac{d\rho_{\rm gw}(f)}{d\ln(f)} \tag{1.76}$$

The associated power spectral density can be written [57]

$$S_{\rm GW} = \frac{3H_0^2}{10\pi^2} f^{-3}\Omega(f). \tag{1.77}$$

Note that, as for the Cosmic Microwave Background Radiation (CMBR), the primordial gravitational waves would be highly redshifted from the expansion of the universe, but likely to a much greater degree, since they would have decoupled from matter at vastly earlier times. A more convenient reformulation in amplitude spectral density can be written as

$$h(f) \equiv [S_{\rm GW}(f)]^{1/2} = (5.6\times10^{-22})h_{100}[\Omega(f)]^{1/2}\left(\frac{100\,{\rm Hz}}{f}\right)^{3/2}\ {\rm Hz}^{-1/2}, \tag{1.78}$$

where $h_{100} \equiv H_0/(100\,{\rm km/s/Mpc})$. A key question is what range of values is expected for $\Omega(f)$. Fig. 1.9 shows a range of expectations vs. frequency. The light-blue curve is a rough estimate from standard inflationary scenarios [58, 59]. This figure also shows direct limits on GW energy density from comparison of observed abundances of elements with predictions from Big Bang nucleosynthesis (BBN) [60], in addition to limits derived from measurements of anisotropies in the CMBR [61].

A completely different source of cosmological, isotropic stochastic waves could come from cosmic strings. Fig. 1.9 shows a range of predictions of stochastic radiation energy density. Part of this region can be accessible by the future space-interferometers like LISA. A more conventional source of isotropic, stochastic radiation is the superposition of radiation from many distant events, such as binary coalescences from compact stars too far away to be seen individually [62]. In the terrestrial band these coalescences could be from stellar NS-NS, NS-BH and BH-BH systems. A detailed analysis [63] suggests that this background could be detectable by the advaced detectors and could present a significant background for 3rd-generation searches. In the space-based band ($\sim 10^{-4}$ to 1 Hz for LISA design [15]), the coalescences could be from SuperMassive Black Hole Mergers (SMBHM), e.g. from galaxy collisions.

Chapter 2

Continuous Gravitational Waves

LIGO's detection in September 2015 of GWs from the coalescence of heavy stellar-mass black holes (GW150914) [3] marked the beginning of observational gravitational wave astronomy. As discussed in Sec. 1.5, this event falls under the distinctive class of sources known as CBCs. Another significant type of GW source is continuous waves (CWs), emitted by compact spinning objects, most notably non-axisymmetric neutron stars within our galaxy. Although these sources are relatively closer (approximately several kpc compared to tens to hundreds of Mpc), they are expected to produce GW strain amplitudes that are orders of magnitude weaker than those from CBCs, around $\mathcal{O}(10^{-25})$ or smaller, compared to $\mathcal{O}(10^{-21})$. Detecting such weak signals relies on integrating data over extended periods, but this approach typically incurs substantial computational costs due to the need to finely cover a large parameter space (e.g., frequency evolution, sky location, and potential orbital parameters). In nearly every CW search, the achievable sensitivity is constrained by finite computational resources.

This chapter addresses the CW radiation detectable by current-generation ground-based interferometers. It covers the phenomenology of CW radiation, provides a brief overview of the neutron star population and potential detections, and describes the standard physical model for waveform characterization. Additionally, the chapter outlines the analysis methods for detecting CWs, detailing both coherent methods, which analyze the entire observation data set as a whole, and incoherent methods, which divide the observation period into N segments for individual analysis before combining the results.

2.1 Potential sources of CWs

In the frequency band of current ground-based detectors, the primary sources of CWs are galactic, non-axisymmetric neutron stars spinning fast enough that their rotation frequencies fall within the LIGO and Virgo detectable range. These nearby neutron stars provide a "conventional" source of CW radiation, as astrophysically extreme as these objects are. An even more exotic proposed source is a "cloud" of bosons, such as QCD axions, surrounding a fast-spinning black hole. These bosons can condense into a few energy levels in enormous numbers, allowing for coherent GW emission from boson annihilation

or level transitions. This chapter focuses primarily on conventional neutron stars, but the exotic boson cloud scenario is also briefly discussed.

2.1.1 Fast-spinning Neutron Stars

Spinning neutron stars could generate detectable GWs through several possible mechanisms. Isolated neutron stars may exhibit intrinsic non-axisymmetry due to residual crustal deformation (e.g., from cooling and cracking of the crust) [64], from a non-axisymmetric distribution of magnetic field energy trapped beneath the crust [65], or from a pinned neutron superfluid component in the star's interior [66–68]. The maximum allowed asymmetries depend on the neutron star equation of state [69] and the breaking strain of the crust [70].

Normal modes of oscillations can also arise, including r-modes in which quadrupole mass currents emit GWs [71]. However, it has been observed [72] that the detectability of emitted radiation from young isolated neutron stars, where mode saturation appears to occur at low r-mode amplitudes due to various dissipative effects, might be doubtful. A recent study [73] is more optimistic about newborn neutron stars. Conversely, the same authors find that r-mode emission from millisecond pulsars is likely to be undetectable by Advanced LIGO [74]. The concept of a runaway rotational instability was first identified for high-frequency f-modes [75], known as the Chandrasekhar-Friedman-Schutz instability. However, realistic viscosity effects seem likely to suppress this effect in conventional neutron star production [76]. The stability of f-modes could play a significant role, however, for a supermassive or hypermassive neutron star formed as the remnant of a binary neutron star merger [77].

Furthermore, binary neutron stars might exhibit direct non-axisymmetry due to non-isotropic accretion [78], a phenomenon also possible for isolated young neutron stars that have experienced fallback accretion shortly after their formation. Additionally, r-modes induced by accretion spin-up can contribute to this effect. The detection of CWs generated by these spinning neutron stars can provide significant insights into neutron star structure and the equation of state of nuclear matter under extreme pressures, particularly when combined with electromagnetic observations of the same star.

It is estimated that there are $\mathcal{O}(10^{8-9})$ neutron stars in our galaxy [79], yet only about 2500 have been detected, mostly as radio pulsars. This small detection fraction is expected due to several factors. Radio pulsations require high magnetic fields and rotation frequencies. An early study [80] suggested the relationship $B \cdot f_{\mathrm{rot}}^2 > 1.7 \times 10^{11}\,\mathrm{G} \times (\mathrm{Hz})^2$, based on a model of radiation dominated by electron-positron pair creation in the stellar magnetosphere. While this model aligns broadly with empirical observations, the resulting "death line" in the period and period derivative plane is perhaps better understood as a "valley" [81]. As a result, isolated pulsars have pulsation lifetimes of approximately $\mathcal{O}(10^7$ years) [82], after which they effectively become radio-invisible. Over this timescale, they also cool to a point where thermal X-ray emission becomes difficult to detect. While X-ray emission from steady accretion of the interstellar medium (ISM) is possible, birth kick velocities significantly suppress such accretion [83], which depends on the inverse cube of the star's velocity through the ISM.

A distinct population of pulsars and non-pulsating neutron stars exists within binary systems. In these binaries, accretion from a non-compact companion star can lead to "recycling," where a spun-down neutron star regains angular momentum from the infalling matter. The rotation frequencies achievable through this spin-up are remarkable—the fastest known rotator is J1748-2446ad at 716 Hz [84]. One notable class of such systems is low mass X-ray binaries (LMXBs), where the neutron star ($\sim 1.4 M_\odot$) has a much lighter companion ($\sim 0.3 M_\odot$) that overfills its Roche lobe, causing material to spill onto an accretion disk surrounding the neutron star or directly onto the star near its magnetic polar caps. As the donor companion star eventually shrinks and decouples from the neutron star, the neutron star can retain a significant portion of its maximum angular momentum and rotational energy. Due to the neutron star's magnetic field decreasing during accretion (through processes that are not fully understood), the spin-down rate after decoupling can be very small. A commonly used phenomenological model for spin-down is a power law

$$\dot{f} = K f^n, \tag{2.1}$$

where f is the star's instantaneous frequency (rotational $f_{\rm rot}$ or gravitational: $f_{\rm GW} \propto f_{\rm rot}$), $\dot{f}$ is the first time derivative of f, and K is a negative constant for most stars. However, there are exceptions for stars experiencing significant acceleration toward us due to being near a deep gravitational well, such as in the core of a globular cluster. The exponent n, known as the *braking index*, varies depending on the spin-down mechanism. The four most commonly discussed theoretical braking indices are:

- $n = 1$ – "Pulsar wind" (extreme model)
- $n = 3$ – Magnetic dipole radiation
- $n = 5$ – Gravitational mass quadrupole radiation ("mountain")
- $n = 7$ – Gravitational mass current quadrupole radiation (r-modes)

While other oscillation modes that generate gravitational waves are possible, the $n = 5$ and $n = 7$ modes are considered the most promising. Assuming this power law has been applicable since the star's birth, the age τ of the star can be related to its birth rotation frequency f_0 and its current frequency f as follows:

$$\tau = -\frac{f}{(n-1)\dot{f}} \left[1 - \left(\frac{f}{f_0} \right)^{(n-1)} \right], \tag{2.2}$$

and if $f \ll f_0$,

$$\tau \approx -\frac{f}{(n-1)\dot{f}}. \tag{2.3}$$

In radio pulsar astronomy, a common assumption is a braking index of $n = 3$, which leads to the nominal magnetic dipole age of a star being defined as

$$\tau_{\rm mag} = -\frac{f}{2\dot{f}}. \tag{2.4}$$

From the general power-law spin-down model given in Eq. (2.1), the second frequency derivative can be expressed as

$$\ddot{f} = nKf^{n-1}\dot{f} = nK^2 f^{2n-1}, \tag{2.5}$$

which allows the current braking index to be determined if the second time derivative of the spin frequency can be measured accurately

$$n = \frac{f\ddot{f}}{\dot{f}^2}. \tag{2.6}$$

2.1.1.1 Pulsar wind $n = 1$

In the early days of the field of pulsar astronomy [85], it was recognized that the outflow of relativistic particles (mainly electrons and positrons, with some ions) from the magnetosphere of a fast-spinning neutron star could induce a spin-down torque. This torque could, in principle, rival that from magnetic dipole radiation, as well as distort the shape of the magnetic field lines and affect the dipole radiation [86]. In this somewhat simplistic model, the spin-down is primarily driven by a braking torque from a return current (predominantly counter-flowing electrons and positrons) crossing magnetic field lines in the star's polar cap regions [87], leading to a braking index $n = 1$. A more recent study on magnetar spin-down [88] proposed a model where sporadic high winds follow bursts, with magnetic dipole emission dominating spin-down between bursts. In steady-state, however, considering the interaction between the magnetic field and the magnetospheric plasma, both magnetic dipole emission and pulsar wind contributions tend to produce a braking index of about three [89], which is discussed next. A phenomenological model [90], a variant of the vacuum dipole model featuring an inner magnetosphere strongly coupled to the star, successfully accounts for the braking indices of the Crab and other young pulsars with $n < 1$.

2.1.1.2 Magnetic dipole $n = 3$.

The radiation energy loss due to a rotating magnetic dipole moment is given by [91]

$$\left(\frac{dE}{dt}\right)_{\text{mag}} = -\frac{\mu_0 M_\perp^2 \omega_{\text{rot}}^4}{6\pi c^3}, \tag{2.7}$$

where ω_{rot} is the rotational angular speed and $M_\perp$ is the component of the star's magnetic dipole moment perpendicular to the rotation axis (assumed to be the z-axis): $M_\perp = M\sin(\alpha)$, with α being the angle between the rotation axis and the magnetic pole.

In a pure dipole moment model, the magnetic field strength at the pole on the star's surface is $B_0 = \mu_0 M/2\pi R^3$. By equating the radiation energy loss to the (Newtonian) rotational energy loss $\frac{1}{2}I_{zz}\omega_{\text{rot}}^2$, we obtain

$$\dot{\omega}_{\text{rot}} = -\frac{2\pi}{3}\frac{R^6}{\mu_0 c^3 I_{zz}} B_\perp^2 \omega_{\text{rot}}^3. \tag{2.8}$$

Thus, the magnetic dipole spin-down rate is proportional to the square of $B_\perp = B_0\sin(\alpha)$ and to the cube of the rotation frequency, resulting in a braking index of $n = 3$.

2.1.1.3 Gravitational mass quadrupole (mountain), $n = 5$.

Let's now consider the gravitational radiation expected from these stars. A star's mass quadrupole asymmetry is commonly characterized by its equatorial ellipticity

$$\epsilon \equiv \frac{|I_{xx} - I_{yy}|}{I_{zz}}. \tag{2.9}$$

An oblate spheroid naturally has a polar ellipticity, but without precession, such deformation does not lead to GW emission. Hereafter, "ellipticity" will refer to equatorial ellipticity, often attributed to a "mountain." For a star at a distance d and spinning about its approximate symmetry axis of rotation (z, assumed to be optimally aligned towards Earth), the expected intrinsic strain amplitude h_0 is

$$\begin{aligned} h_0 &= \frac{4\pi^2 G \epsilon I_{zz} f_{GW}^2}{c^4 d} \\ &= (1.1 \times 10^{-24}) \left(\frac{\epsilon}{10^{-6}}\right) \left(\frac{I_{zz}}{I_0}\right) \left(\frac{f_{GW}}{1\text{ kHz}}\right)^2 \left(\frac{1\text{ kpc}}{d}\right), \end{aligned} \tag{2.10}$$

where $I_0 = 10^{38}$ kg·m^2(10^{45} g·cm^2) is a nominal moment of inertia for a neutron star, and the gravitational radiation is emitted at $f_{GW} = 2f_{rot}$. The total power emitted in gravitational waves from the star (integrated over all angles) is

$$\begin{aligned} \frac{dE}{dt} &= -\frac{32G}{5c^5} I_{zz}^2 \epsilon^2 \omega_{rot}^6 \\ &= (-1.7 \times 10^{33}\text{ J/s}) \left(\frac{I_{zz}}{I_0}\right)^2 \left(\frac{\epsilon}{10^{-6}}\right)^2 \left(\frac{f_{GW}}{1\text{ kHz}}\right)^6. \end{aligned} \tag{2.11}$$

Equating this loss to the reduction of rotational kinetic energy $\frac{1}{2} I_{zz} \omega_{rot}^2$ leads to the spin-down relationship

$$\begin{aligned} \dot{f}_{GW} &= -\frac{32\pi^4 G}{5c^5} I_{zz} \epsilon^2 f_{GW}^5 \\ &= (-1.7 \times 10^{-9}\text{ Hz/s}) \left(\frac{\epsilon}{10^{-6}}\right)^2 \left(\frac{f_{GW}}{1\text{ kHz}}\right)^5, \end{aligned} \tag{2.12}$$

highlighting the braking index of 5. For an observed neutron star with measured f and $\dot{f}$, the "spin-down limit" on the maximum allowed strain amplitude can be defined by equating the power loss in Eq. (2.11) to the time derivative of the rotational kinetic energy $\frac{1}{2} I_{zz} \omega_{rot}^2$

$$\begin{aligned} h_{\text{spin-down}} &= \frac{1}{d} \sqrt{-\frac{5GI_{zz}}{2c^3} \frac{\dot{f}_{GW}}{f_{GW}}} \\ &= (2.6 \times 10^{-25}) \left[\frac{1\text{kpc}}{d}\right] \left[\left(\frac{1\text{ kHz}}{f_{GW}}\right) \left(\frac{-\dot{f}_{GW}}{10^{-10}\text{ Hz/s}}\right) \left(\frac{I_{zz}}{I_0}\right)\right]^{1/2}. \end{aligned} \tag{2.13}$$

Thus, for each observed pulsar with a measured frequency, spin-down rate, and distance d, one can determine whether energy conservation permits the detection of GWs in an optimistic scenario. Unfortunately, nearly all known pulsars have spin-down strain limits below the detection capabilities of the LIGO and Virgo detectors at their current sensitivities.

2.1.1.4 Gravitational mass current quadrupole (r-modes) $n = 7$.

Different frequency scalings apply to mass quadrupole and mass current quadrupole emissions. The most promising source of mass current non-axisymmetry in neutron stars is believed to be "r-modes," which result from the fluid motion of neutrons (or protons) in the crust or core of the star. These currents, much like Rossby waves in Earth's atmosphere, are deflected by Coriolis forces, creating spatial oscillations [92]. These r-modes can become inherently unstable, arising from azimuthal interior currents that are retrograde in the star's rotating frame but appear prograde in an external reference frame. Consequently, the quadrupolar gravitational wave emission due to these currents amplifies the current's strength, leading to a positive-feedback loop and a potential intrinsic Chandrasekhar-Friedman-Schutz [93] instability. The frequency of such emissions is expected to be at most approximately 4/3 of the rotation frequency [94, 95]. Using Owen's notation [95], the mass current can be modeled as a velocity field perturbation δv_j, which, when integrated, gives the intrinsic strain amplitude observed at a distance d

$$\begin{aligned} h_0 &= \sqrt{\frac{512\pi^7}{5}}\frac{G}{c^5 d} f_{\rm GW}^3 \alpha M R^3 \tilde{J} \\ &= 3.6\times10^{-26}\left(\frac{1\ \text{kpc}}{d}\right)\left(\frac{f_{\rm GW}}{100\ \text{Hz}}\right)^3\left(\frac{\alpha}{10^{-3}}\right)\left(\frac{R}{11.7\ \text{km}}\right)^3. \end{aligned} \tag{2.14}$$

Here, α represents the dimensionless r-mode amplitude, M is the stellar mass, R is the radius, and $\tilde{J}$ is a dimensionless function of the stellar equation of state. For a Newtonian polytrope with an index of 1, $\tilde{J} \approx 0.0164$ [95], as used in the fiducial Eq. (2.14). The energy loss in this model is given by [96]

$$\frac{dE}{dt} = -\frac{1024\pi^9}{25}\frac{G}{c^7} f_{\rm GW}^8 \alpha^2 M^2 R^6 \tilde{J}^2. \tag{2.15}$$

Equating this loss to the reduction in rotational kinetic energy $\frac{1}{2} I_{zz} \omega_{\rm rot}^2$ results in the spin-down relationship

$$\begin{aligned} \dot{f}_{\rm GW} &= -\frac{4096\pi^7}{225}\frac{G M^2 R^6 \tilde{J}^2}{c^7 I_{zz}} \alpha^2 f_{\rm GW}^7 \\ &= -9.0\times10^{-14}\ \text{Hz/s}\left(\frac{R}{11.7\ \text{km}}\right)^6\left(\frac{\alpha}{10^{-3}}\right)^2\left(\frac{f_{\rm GW}}{100\ \text{Hz}}\right)^7, \end{aligned} \tag{2.16}$$

where the braking index of 7 is evident. As before, one can define a spin-down limit based on pure r-mode radiation

$$h_{\text{spin-down}} = \frac{1}{d}\sqrt{-\frac{45G}{8c^3 I_{zz}}\frac{\dot{f}_{\rm GW}}{f_{\rm GW}}}, \tag{2.17}$$

where the ratio of this spin-down limit to the one given in Eq. (2.13) is 3/2, arising from the different ratios of GW signal frequency to spin frequency for mass quadrupole versus mass current quadrupole radiation.

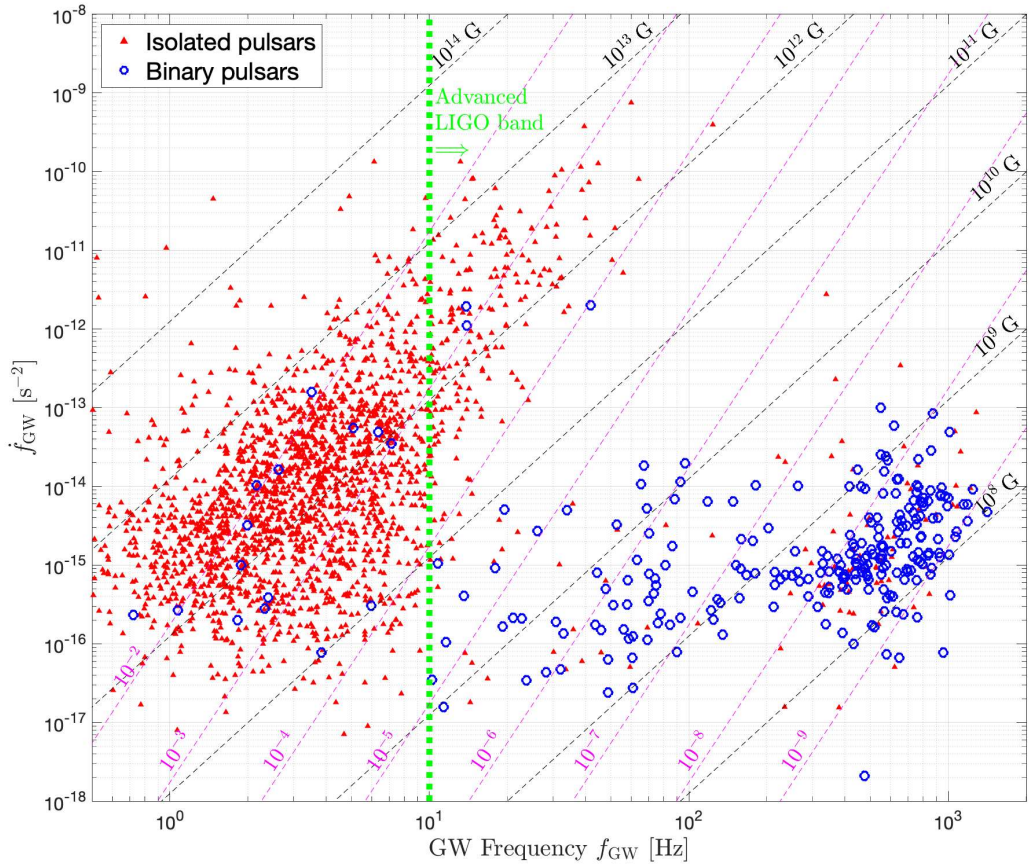


FIGURE 2.1: Nominal expected GW frequencies and frequency derivatives for known pulsars, taken from [99]. Closed triangles denote isolated stars, while open circles represent binary stars. Contours indicate constant magnetic fields (ellipticities) for spin-down dominated by magnetic dipole and gravitational mass quadrupole emissions. The vertical dotted line marks the approximate sensitivity band for Advanced LIGO at design sensitivity. A similar sensitivity band applies to the design sensitivities of the Advanced Virgo and KAGRA detectors. [98].

2.1.1.5 Neutron star survey.

Gravitars refer to neutron stars with spin-down dominated by GW energy loss [97]. Although it is believed that most known pulsars are not gravitars, the model is still useful in setting bounds on what might be detectable. Fig. 2.1 displays a subset of pulsars in the $f_{\rm GW} - \dot{f}_{\rm GW}$ plane, assuming $f_{\rm GW} = 2f_{\rm rot}$. Isolated stars and binary stars are denoted by closed circles and open triangles, respectively. A vertical dashed line marks the approximate detection bandwidth for Advanced LIGO at design sensitivity (~ 10 Hz and above). This frequency boundary also applies to the design sensitivities of the Advanced Virgo and KAGRA detectors [98]. Contours indicate constant magnetic field, assuming spin-down is dominated by magnetic dipole emission ($n = 3$). Additionally, higher slope contours represent constant ellipticity. An interesting deficit of millisecond pulsars with extremely low period derivatives appears consistent with a population of sources having a minimum ellipticity of about $\sim 10^{-9}$ and additional spin-down losses from magnetic dipole radiation, as indicated by the near absence of sources to the right of the $\epsilon = 10^{-9}$ line in Fig. 2.1. Conversely, lower-frequency, younger pulsars with high spin-downs are observed, with the highest being 7.6×10^{-10} Hz/s (Crab pulsar).

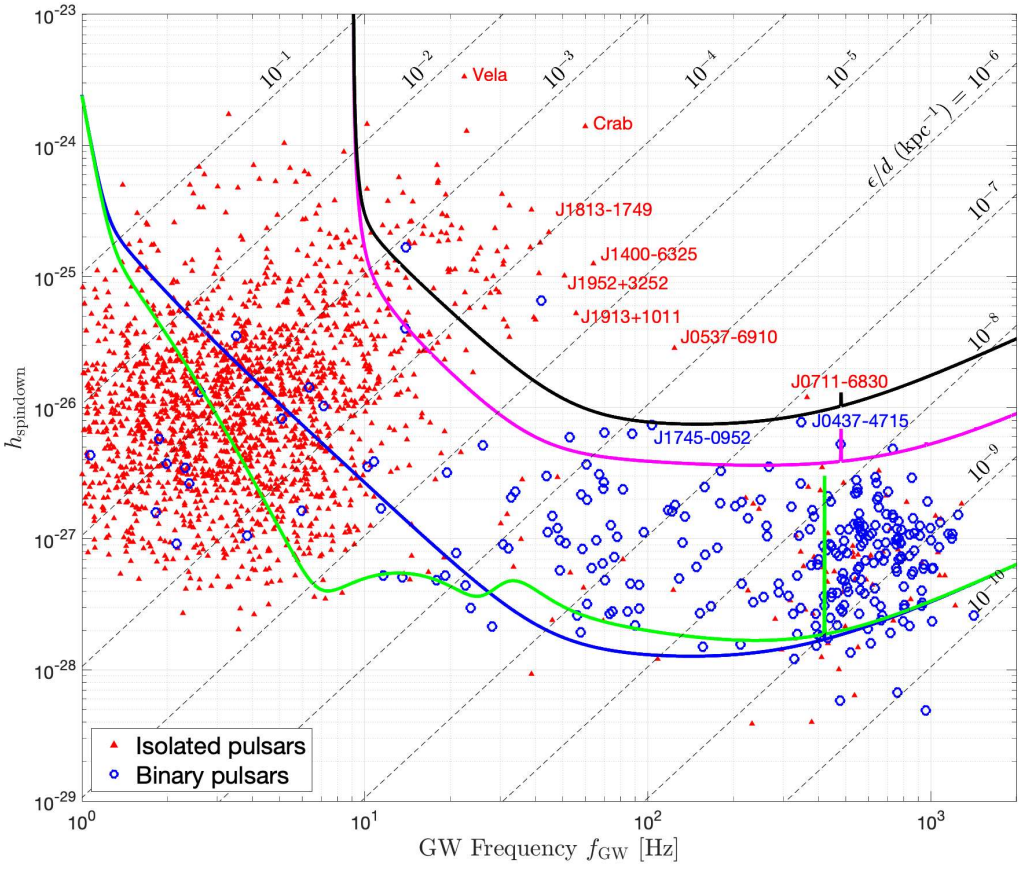


FIGURE 2.2: Nominal expected GW frequencies and nominal strain spin-down limits for known pulsars, taken from [99]. Closed triangles indicate isolated stars. Closed triangles represent isolated stars, while open circles indicate binary stars. Solid curves denote the nominal strain noise sensitivity for the O3 observing run (black), as well as the anticipated sensitivities for a 2-year advanced detector run at design sensitivity (magenta) and a 5-year Einstein Telescope run with two different designs: ETB (blue) and ETC (green). Dashed diagonal lines show particular ratios of ellipticity to distance. Additionally, a subset of notable pulsars are labeled on the plot.

Using Eq. (2.13), these known pulsars can be mapped onto a plane of $f_{\mathrm{GW}} - h_0$ under the gravitar assumption, as depicted in Fig. 2.2. The spin-down strain limit (for $n = 5$) is shown on the vertical axis. Additionally, corresponding contours of constant implied values of ϵ/d, where d is the distance to the star, are indicated under the gravitar assumption. Detector network sensitivities for advanced detectors at design sensitivity [98] and for two proposed configurations of the "3rd-generation" Einstein Telescope (ET) [100] are also shown. Another 3rd-generation proposal, the "Cosmic Explorer" [101], would have performance comparable to that of the ET, being more sensitive at frequencies above ~ 10 Hz and less sensitive at lower frequencies.

In Fig. 2.2, the sensitivities of "advanced detectors" are represented by calculations for two Advanced LIGO detectors operating continuously over two observing years, referred to as the "O4/O5 run." Although the O4 run, has already started at the beginning of 2023, and will likely last only about a year and a half [98], it may not fully achieve the original Advanced LIGO design sensitivity, the subsequent O5 run in the "A+" configuration is expected to significantly exceed Advanced LIGO sensitivity and last

for over a year. This makes the detector sensitivities assumed here conservative in principle. Including Advanced Virgo and KAGRA in the network sensitivity would further improve these sensitivities. However, the O4/O5 observing time assumed here does not account for realistic dead-time losses, which can be substantial ($\sim 25\%$ per detector [102]). The detection sensitivities illustrated in Fig. 2.2 assume a *targeted search* using known pulsar ephemerides. If a star is marked above a sensitivity curve, detecting it is theoretically possible if its spin-down classifies it as a gravitar. It is important to note that Eq. (2.13) has been applied with a nominal moment of inertia I_{zz}, but the uncertainty in I_{zz} is approximately a factor of two, depending on the equation of state and stellar mass [103].

2.1.2 Axion clouds bound to black holes

In recent years, a compelling hypothesis has gained traction: dark matter might not only be composed of electromagnetically invisible massive bosons, such as axions, but these bosons could be disproportionately concentrated around rapidly spinning black holes [104]. These bosons could potentially be created spontaneously through the extraction of energy from the black hole's rotation [105], forming a Bose-Einstein "cloud" where most quanta occupy a small number of energy levels. For a cloud bound to a black hole, the approximate inverse-square law attraction outside the Schwarzschild radius leads to energy level spacing similar to that of a hydrogen atom. The phenomenon of superradiance near a rapidly spinning black hole (with angular momentum close to the maximum allowed by General Relativity) can significantly amplify the number of quanta in the low-lying levels. Bosons in non-s ($\ell > 0$) negative-energy states can be thought of as moving in a potential well formed by an ℓ-dependent centrifugal barrier at $r > r_{\text{Schwarzschild}}$ and a potential that rises towards zero as $r \to \infty$. Wave function penetration into the black hole ergosphere enables energy transfer from the black hole's spin [106], facilitating the creation of new quanta.

In the axion scenario, two particular gravitational wave emission modes of interest can arise, both potentially leading to intense coherent radiation [107]. In one mode, axions can annihilate each other to produce gravitons, with the frequency of the graviton being twice that of the axion mass: $f_{\text{graviton}} = 2m_{\text{axion}}c^2/h$. In another mode, emission occurs from level transitions of quanta within the cloud. This Bose condensation is most pronounced when the reduced Compton wavelength of the axion is comparable to but larger than the Schwarzschild radius of the black hole:

$$\frac{\lambda}{2\pi} = \frac{\hbar}{m_{\text{axion}}c} \gtrsim \frac{2GM_{\text{BH}}}{c^2} \Rightarrow m_{\text{axion}} \lesssim (7 \times 10^{-11}\ \text{eV}/c^2)\frac{M_\odot}{M_{\text{BH}}}, \tag{2.18}$$

where $\hbar$ and G are Planck's and Newton's constants respectively. Given the vast uncertainty in axion masses that could account for dark matter [108], the relatively narrow mass window accessible to currently feasible CW searches ($\sim$ 2 orders of magnitude) makes the search for such emissions a quintessential example of "almost" physics, where one can only hope that nature places the axion within this illuminated area of a vast parameter space.

Searching for these potential CW sources generally does not require a fundamental modification of the search methods, but they can be optimized for the specific frequency bands defined by annihilation and level transition emissions. Directed searches targeting known black hole locations can achieve better sensitivity than all-sky searches. However, in the context of string axiverse models, the axion cloud [104] can undergo significant self-interactions, resulting in substantial frequency evolution of the signal and uncertainty in that evolution, which is a lesser issue for the proposed QCD axion [107]. An intriguing possibility is that a black hole formed from a detected BBH merger could serve as a natural target for follow-up CW searches [109]. So far, no published searches have been specifically designed for a black hole axion cloud source, but existing limits on neutron star emissions, even if not optimized for this purpose, can be reinterpreted to constrain such emissions [107]. Recent research [110] suggests that the absence of a detectable stochastic gravitational wave background from extragalactic black holes already places significant constraints on axion masses relevant to CW searches.

2.2 Continuous Wave signal model

We begin the analysis with the beam-pattern functions F_+ and $F_\times$, as previously shown in Eq. (1.43). Let's start by rephrasing Eq. (1.41) in an alternative form

$$h(t) = \frac{1}{2}\boldsymbol{n_1} \cdot \tilde{H}(t)\boldsymbol{n_1} - \frac{1}{2}\boldsymbol{n_2} \cdot \tilde{H}(t)\boldsymbol{n_2}, \tag{2.19}$$

where $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ are the unit vectors parallel to the arms 1 and 2 of the Michelson interferometer, respectively. The term $\tilde{H}$ represents the spatial-metric perturbation caused by the wave in the detector's proper reference frame, with the dot indicating the scalar product. The matrix $\tilde{H}$ is defined as

$$\tilde{H}(t) = M(t)H(t)M(t)^T, \tag{2.20}$$

where M is a 3-dimensional orthogonal matrix transforming from the Cartesian frame (x_w, y_w, z_w) to the detector's frame (x_d, y_d, z_d). Since CWs are long-lasting, the phase and polarization angle's time dependence is crucial and must be precisely accounted for. Due to the Earth's diurnal motion, the beam-patterns F_+ and $F_\times$ are periodic functions with a period of one sidereal day. Our aim is to explicitly extract this time dependence and express F_+ and $F_\times$ as functions of the right ascension α, declination δ, and polarization angle ψ of the GW source. These angles define the wave reference frame orientation relative to the celestial sphere reference frame. We define a transformation matrix M [111] as

$$M = M_3 M_2 M_1^T, \tag{2.21}$$

where M_1 transforms from wave to celestial sphere coordinates, M_2 transforms from celestial to cardinal coordinates, and M_3 transforms from cardinal to detector coordinates. In celestial sphere coordinates, the z axis aligns with the Earth's rotation axis pointing toward the North pole, the x and y axes lie in the Earth's equatorial plane, and the x axis points toward the vernal point. In cardinal coordinates, the

(x, y) plane is tangent to the Earth's surface at the detector's location, with the x axis in the North-South direction, the y axis in the West-East direction, and the z axis pointing toward zenith. In detector coordinates, the z axis aligns with the cardinal z axis, and the x axis aligns with the first interferometer arm (the y axis aligns with the second arm if the arms are perpendicular). Under these conventions, the matrices M_1, M_2, and M_3 are defined as follows [111]

$$M_1 = \begin{pmatrix} \sin\alpha\cos\psi - \cos\alpha\sin\delta\sin\psi & -\cos\alpha\cos\psi - \sin\alpha\sin\delta\sin\psi & \cos\delta\sin\psi \\ -\sin\alpha\sin\psi - \cos\alpha\sin\delta\cos\psi & \cos\alpha\sin\psi - \sin\alpha\sin\delta\cos\psi & \cos\delta\cos\psi \\ -\cos\alpha\cos\delta & -\sin\alpha\cos\delta & -\sin\delta \end{pmatrix}, \tag{2.22}$$

$$M_2 = \begin{pmatrix} \sin\lambda\cos(\phi_r + \Omega_r t) & \sin\lambda\sin(\phi_r + \Omega_r t) & -\cos\lambda \\ -\sin(\phi_r + \Omega_r t) & \cos(\phi_r + \Omega_r t) & 0 \\ \cos\lambda\cos(\phi_r + \Omega_r t) & \cos\lambda\sin(\phi_r + \Omega_r t) & \sin\lambda \end{pmatrix}, \tag{2.23}$$

$$M_3 = \begin{pmatrix} -\sin(\gamma + \zeta/2) & \cos(\gamma + \zeta/2) & 0 \\ -\cos(\gamma + \zeta/2) & -\sin(\gamma + \zeta/2) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.24}$$

In M_2, λ is the latitude of the detector's site, Ω_r is the Earth's rotational angular velocity, and ϕ_r is a phase defining the Earth's position in its diurnal motion at $t = 0$ (the sum $\phi_r + \Omega_r t$ corresponds to the local sidereal time of the detector's site, i.e., the angle between the local meridian and the vernal point). In M_3, γ defines the orientation of the detector's arms relative to local geographical directions: γ is measured counter-clockwise from East to the bisector of the interferometer arms, and ζ is the angle between the interferometer arms. The vectors $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ from Eq. (2.19) in the detector's reference frame have coordinates

$$\boldsymbol{n}_1 = (1, 0, 0), \quad \boldsymbol{n}_2 = (\cos\zeta, \sin\zeta, 0). \tag{2.25}$$

The values of the angles λ, γ, ζ, and the longitudes L (measured positively westwards) for different detectors are provided in Table 2.1 [112]. To derive explicit formulas for F_+ and $F_\times$, we combine Eqs. (2.19)-(2.24). After extensive algebraic manipulations, we obtain

$$F_+(t) = \sin\zeta\,[a(t)\cos 2\psi + b(t)\sin 2\psi]\,, \tag{2.26}$$

$$F_\times(t) = \sin\zeta\,[b(t)\cos 2\psi - a(t)\sin 2\psi]\,, \tag{2.27}$$

TABLE 2.1: Positions and orientations of detectors.

detector	λ (degrees)	L (degrees)	γ (degrees)	ζ (degrees)
GEO600	52.25	-9.81	68.775	94.33
LIGO Hanford	46.45	119.41	171.8	90
LIGO Livingston	30.56	90.77	243.0	90
VIRGO	43.63	-10.5	116.5	90
TAMA300	35.68	-139.54	225.0	90

where

$$
\begin{aligned}
a(t) = \frac{1}{16}\sin 2\gamma(3-\cos 2\lambda)(3-\cos 2\delta)\cos[2(\alpha-\phi_r-\Omega_r t)] \\
-\frac{1}{4}\cos 2\gamma \sin\lambda(3-\cos 2\delta)\sin[2(\alpha-\phi_r-\Omega_r t)] \\
+\frac{1}{4}\sin 2\gamma \sin 2\lambda \sin 2\delta \cos(\alpha-\phi_r-\Omega_r t) \\
-\frac{1}{2}\cos 2\gamma \cos\lambda \sin 2\delta \sin(\alpha-\phi_r-\Omega_r t) \\
+\frac{3}{4}\sin 2\gamma \cos^2\lambda \cos^2\delta, \qquad (2.28)
\end{aligned}
$$

$$
\begin{aligned}
b(t) = \cos 2\gamma \sin\lambda \sin\delta \cos[2(\alpha-\phi_r-\Omega_r t)] \\
+\frac{1}{4}\sin 2\gamma(3-\cos 2\lambda)\sin\delta \sin[2(\alpha-\phi_r-\Omega_r t)] \\
+\cos 2\gamma \cos\lambda \cos\delta \cos(\alpha-\phi_r-\Omega_r t) \\
+\frac{1}{2}\sin 2\gamma \sin 2\lambda \cos\delta \sin(\alpha-\phi_r-\Omega_r t). \qquad (2.29)
\end{aligned}
$$

Using Eqs. (2.26)-(2.29), the beam-pattern functions can be directly computed for any given time.

2.2.1 The phase of the gravitational-wave signal

In Appendix B.1, the derivation of the time dependence of the phase of the GW signal observed at the detector's location is detailed. The significance of corrections due to the motion of both the detector and the neutron star relative to the solar system barycenter (SSB) reference frame, as well as the importance of relativistic corrections, are considered. Based on the discussion presented in Appendix B.1, the following model for the phase of the GW signal is adopted

$$
\Psi(t) = \Psi_0 + 2\pi \sum_{k=0}^{s} \frac{f_0^{(k)}}{(k+1)!} t^{k+1} + \frac{2\pi}{c}\boldsymbol{n}_0 \cdot \boldsymbol{r}_d(t) \sum_{k=0}^{s} \frac{f_0^{(k)} t^k}{k!}, \qquad (2.30)
$$

where $f_0^{(k)}$ represents the k-th time derivative of the instantaneous frequency evaluated at $t = 0$ at the SSB, $\boldsymbol{n}_0$ is the constant unit vector in the direction of the star in the SSB reference frame, and $\boldsymbol{r}_d$ is the position vector of the detector in that frame. The signal analysis is independent of the number s of spindown parameters, so s is left unspecified. A coordinate system is associated with the SSB reference frame. The z axis of this system is parallel to the x axis of the celestial sphere coordinate system and perpendicular to the ecliptic, coinciding with the Earth's orbital angular momentum vector. In this system, the unit vector $\boldsymbol{n}_0$ pointing towards the star has components

$$
\boldsymbol{n}_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\epsilon & \sin\epsilon \\ 0 & -\sin\epsilon & \cos\epsilon \end{pmatrix} \begin{pmatrix} \cos\alpha\cos\delta \\ \sin\alpha\cos\delta \\ \sin\delta \end{pmatrix}, \qquad (2.31)
$$

where ϵ is the angle between the ecliptic and the Earth's equator. The position vector $\boldsymbol{r}_d$ of the detector in this coordinate system is

$$\boldsymbol{r}_d = R_{\rm ES}\begin{pmatrix}\cos(\phi_o+\Omega_r t)\\ \sin(\phi_o+\Omega_r t)\\ 0\end{pmatrix} + R_E\begin{pmatrix}1 & 0 & 0\\ 0 & \cos\epsilon & \sin\epsilon\\ 0 & -\sin\epsilon & \cos\epsilon\end{pmatrix}\begin{pmatrix}\cos\lambda\cos(\phi_r+\Omega_r t)\\ \cos\lambda\sin(\phi_r+\Omega_r t)\\ \sin\lambda\end{pmatrix}, \tag{2.32}$$

where $R_{\rm ES} = 1\text{AU}$ is the average distance from the Earth's center to the SSB, R_E is the mean radius of the Earth, Ω_0 is the mean orbital angular velocity of the Earth, and ϕ_o is a deterministic phase that defines the Earth's position in its orbital motion at $t = 0$. The eccentricity of the Earth's orbit and the motion of the Earth around the Earth-Moon barycenter are neglected. Substituting Eqs. (2.31) and (2.32) into Eq. (2.30) yields

$$\Psi(t) = \Psi_0 + \Phi(t), \tag{2.33}$$

$$\begin{aligned}\Phi(t) = {} & 2\pi\sum_{k=0}^{s}\frac{f_0^{(k)}}{(k+1)!}t^{k+1}\\ & +\frac{2\pi}{c}\Big[R_{\rm ES}(\cos\alpha\cos\delta\cos(\phi_o+\Omega_r t)+\cos\epsilon(\cos\delta\sin\alpha\cos\delta\sin\alpha+\sin\delta)\sin(\phi_o+\Omega_r t)).\\ & +R_E\{\sin\lambda\sin\delta+\cos\lambda\cos\delta\cos(\alpha-\phi_r-\Omega_r t)\}\Big]\sum_{k=0}^{s}\frac{f_0^{(k)}t^k}{k!}.\end{aligned} \tag{2.34}$$

2.2.2 Wave polarization functions

The following two-component model of the gravitational-wave signal is used

$$h(t) = h_1(t) + h_2(t), \tag{2.35}$$

where the components are defined by

$$h_1(t) = F_+(t)h_{1+}(t) + F_\times(t)h_{1\times}(t), \qquad h_2(t) = F_+(t)h_{2+}(t) + F_\times(t)h_{2\times}(t), \tag{2.36}$$

$$h_{1+}(t) = \frac{1}{8}h_0\sin 2\theta\sin 2\iota\cos\Psi(t), \qquad h_{2+}(t) = \frac{1}{2}h_0\sin^2\theta(1+\cos^2\iota)\cos 2\Psi(t), \tag{2.37}$$

$$h_{1\times}(t) = \frac{1}{4}h_0\sin 2\theta\sin\iota\sin\Psi(t), \qquad h_{2\times}(t) = h_0\sin^2\theta\cos\iota\sin 2\Psi(t). \tag{2.38}$$

The beam-pattern functions F_+, $F_\times$ are given by Eqs. (2.26)-(2.29) and the phase Ψ is given by Eqs. (2.33) and (2.34). The model of the signal defined by Eqs. (2.35)-(2.38) encapsulates the quadrupole gravitational waves emitted by a freely precessing axisymmetric star. The angle θ, known as the wobble angle, is the angle between the total angular momentum vector of the star and the star's axis of symmetry. The angle ι is the angle between the total angular momentum vector of the star and the line of sight from the star to the Earth. The amplitude h_0 is determined by

$$h_0 = \frac{16\pi^2 G\epsilon I f^2}{c^4 r}, \tag{2.39}$$

where f is the sum of the frequency of rotation of the star and the frequency of precession, I denotes the moment of inertia concerning the rotation axis, ϵ represents the poloidal ellipticity of the star, and r is the distance to the star. For small wobble angles, the signal h_1 predominates. Comprehensive details of the model are discussed in [113]. When $\theta = \pi/2$ the h_1 component vanishes, and for this special case, the h_2 component represents the quadrupole wave from a triaxial ellipsoid rotating about a principal axis with frequency f. Under these conditions, the amplitude h_0 is recalculated using the ellipticity defined by

$$\epsilon = \frac{I_1 - I_2}{I}, \tag{2.40}$$

where I_1 and I_2 are the moments of inertia of the star with respect to the principal axes orthogonal to the rotation axis. This model is further explored in [51].

After replacing physical constants in Eq. (2.39) with their numerical values, the amplitude expression simplifies to

$$h_0 = 4.23 \times 10^{-25} d_o \left(\frac{f}{100\ \text{Hz}}\right)^2, \tag{2.41}$$

where

$$d_o := \left(\frac{\epsilon}{10^{-5}}\right)\left(\frac{I}{10^{45}\ \text{g cm}^2}\right)\left(\frac{1\ \text{kpc}}{r}\right). \tag{2.42}$$

By integrating the beam-pattern functions from Eqs. (2.26) the signal described by Eqs. (2.35)-(2.38) can be expressed as

$$h(t) = \sum_{i=1}^{4} A_{1i} h_{1i}(t) + \sum_{i=1}^{4} A_{2i} h_{2i}(t), \tag{2.43}$$

where the eight amplitudes A_{1i} and A_{2i} are given by

$$A_{11} = h_0 \sin\zeta \sin 2\theta \left[\frac{1}{8}\sin 2\iota \cos 2\psi \cos\Phi_0 - \frac{1}{4}\sin\iota \sin 2\psi \sin\Phi_0\right], \tag{2.44}$$

$$A_{12} = h_0 \sin\zeta \sin 2\theta \left[\frac{1}{4}\sin\iota \cos 2\psi \sin\Phi_0 + \frac{1}{8}\sin 2\iota \cos 2\psi \cos\Phi_0\right], \tag{2.45}$$

$$A_{13} = h_0 \sin\zeta \sin 2\theta \left[-\frac{1}{8}\sin 2\iota \cos 2\psi \sin\Phi_0 - \frac{1}{4}\sin\iota \sin 2\psi \cos\Phi_0\right], \tag{2.46}$$

$$A_{14} = h_0 \sin\zeta \sin 2\theta \left[\frac{1}{4}\sin\iota \cos 2\psi \cos\Phi_0 - \frac{1}{8}\sin 2\iota \sin 2\psi \sin\Phi_0\right], \tag{2.47}$$

$$A_{21} = h_0 \sin\zeta \sin^2\theta \left[\frac{1}{2}(1+\cos^2\iota)\cos 2\psi \cos 2\Phi_0 - \cos\iota \sin 2\psi \sin 2\Phi_0\right], \tag{2.48}$$

$$A_{22} = h_0 \sin\zeta \sin^2\theta \left[\frac{1}{2}(1+\cos^2\iota)\sin 2\psi \cos 2\Phi_0 + \cos\iota \cos 2\psi \sin 2\Phi_0\right], \tag{2.49}$$

$$A_{23} = h_0 \sin\zeta \sin^2\theta \left[-\frac{1}{2}(1+\cos^2\iota)\cos 2\psi \sin 2\Phi_0 - \cos\iota \sin 2\psi \cos 2\Phi_0\right], \tag{2.50}$$

$$A_{24} = h_0 \sin\zeta \sin^2\theta \left[-\frac{1}{2}(1+\cos^2\iota)\sin 2\psi \sin 2\Phi_0 + \cos\iota \cos 2\psi \cos 2\Phi_0\right]. \tag{2.51}$$

The amplitudes A_{1i} and A_{2i} depend on the parameters h_0, θ, ψ, ι, and Φ_0 along with the angle ζ. The time dependent functions h_{1i} have the form

$$h_{11} = a(t)\cos l\Phi(t), \qquad h_{12} = b(t)\cos l\Phi(t), \tag{2.52}$$

$$h_{13} = a(t)\sin l\Phi(t), \qquad h_{14} = b(t)\sin l\Phi(t), \quad l = 1, 2. \tag{2.53}$$

where the functions a and b are given by Eqs. (2.28) and (2.29), respectively, and Φ is the phase given by Eq. (2.34). The modulation amplitudes a and b depend on the right ascension α and the declination δ of the source (they also depend on the angles λ and γ). The phase Φ depends on the frequency f_0, spin-down parameters $f_0^{(k)}(k = 1, \ldots, s)$, and on the angles α, δ. We call the parameters f_0, $f_0^{(k)}$, and the angles α, δ the *phase parameters*. Moreover the phase Φ depends on the latitude λ of the detector. The whole signal h depends on $8 + s$ unknown parameters: h_0, θ, ψ, ι, Φ_0, α, δ, f_0, $f_0^{(k)}$.

The frequency domain characteristics of the GW signal are notable for their complexity. The signal includes two primary components with carrier frequencies f_0 and $2f_0$ both subject to extensive amplitude and frequency modulation. The amplitude modulation creates five distinct frequency lines for each component, $f_0 - 2f_r$, $f_0 - f_r$, f_0, $f_0 + f_r$, $f_0 + 2f_r$, where f_r is the frequency of rotation of Earth ($f_r \simeq 10^{-5}$ Hz) and the same for frequency $2f_0$. while the frequency modulation causes line broadening, significantly affecting the detectable signal profile during extended observations.

2.2.3 Optimal filter for the amplitude modulated signal

2.2.3.1 Maximum likelihood detection

The signal given by Eq. (2.43) will be buried in the noise of a detector. Thus, the problem is detecting the signal and estimating its parameters. A standard method is the *maximum likelihood detection* which involves maximizing the likelihood function Λ with respect to the signal parameters. If the maximum of Λ exceeds a certain threshold, calculated based on the acceptable false alarm probability, the signal is considered detected. The parameter values that maximize Λ are termed the *maximum likelihood estimators* of the signal parameters. The magnitude of the maximum of Λ indicates the probability of detecting the signal. Assume the noise n in the detector is an additive, stationary, Gaussian, and zero-mean continuous random process. Then the data x (if the signal h is present) can be written as

$$x(t) = n(t) + h(t). \tag{2.54}$$

The log likelihood function is

$$\log \Lambda = (x|h) - \frac{1}{2}(h|h), \tag{2.55}$$

where the scalar product $(\cdot|\cdot)$ is defined by

$$(x|y) := 4\Re\left\{\int_0^\infty \frac{\tilde{x}(f)\tilde{y}^*(f)}{S_h(f)} df\right\}, \tag{2.56}$$

where S_h is the *one-sided spectral density* of the detector's noise. The GW signal given by Eq. (2.43) consists of two narrow-band components around the frequencies f_0 and $2f_0$ and thus, to a very good accuracy, the likelihood ratio is given by

$$\log \Lambda \cong (x|h_1) - \frac{1}{2}(h_1|h_1) + (x|h_2) - \frac{1}{2}(h_2|h_2). \tag{2.57}$$

This suggests considering the two components of the response function (2.43) as two independent signals. Taking the first component h_1 of the signal, one can assume that over the bandwidth of the signal $S_h(f)$ is nearly constant and equal to $S_h(f_0)$ where f_0 is the frequency of the signal h_1 at $t = 0$. Thus, the scalar product can be approximated by

$$(x|h_1) \cong \frac{2}{S_h(f_0)} \int_{-T_0/2}^{T_0/2} x(t)h_1(t)dt, \tag{2.58}$$

where T_0 is the observation time and where the observation interval is $[-T_0/2, T_0/2]$. Introducing the following scalar product is useful

$$(x||y) := \frac{2}{T_0} \int_{-T_0/2}^{T_0/2} x(t)y(t)dt. \tag{2.59}$$

As long as the detector's noise is stationary over the observation period, this is a valid scalar product. In realistic observations, the detector's noise will vary slowly during this period, however, this issue is not addressed here. The log likelihood function for this signal is approximately given by

$$\log \Lambda_1 \cong \frac{T_o}{S_h(f_0)} \left[(x||h_1) - \frac{1}{2}(h_1||h_1) \right]. \tag{2.60}$$

The maximum likelihood estimators can be found by maximizing the following *normalized log likelihood function*

$$\log \Lambda_1' = (x||h_1) - \frac{1}{2}(h_1||h_1). \tag{2.61}$$

The normalized log likelihood function does not involve explicitly the spectral density of the noise in the detector and the signal h_1 depends linearly on four amplitudes A_{1i}. The amplitudes depend on the five unknown parameters h_0, θ, ψ, ι, and Φ_0 and are independent. The likelihood equations for the amplitudes A_{1i} are given by

$$\frac{\partial \ln \Lambda_1'}{\partial A_{1i}} = 0, \quad i = 1, \ldots, 4. \tag{2.62}$$

In this case, the above set of equations is equivalent to the following set of linear algebraic equations

$$\sum_{j=1}^{4} M_{ij} A_{1j} = (x||h_{1i}), \quad i = 1, \ldots, 4, \tag{2.63}$$

where the components M_{ij} of the 4×4 matrix $\mathcal{M}$ are given by

$$M_{ij} := (h_{1i}||h_{1j}). \tag{2.64}$$

Since over a typical observation time T_0, the phase Φ will have many oscillations, to a very good accuracy,

$$(h_{11}||h_{13}) \cong 0, \quad (h_{11}||h_{14}) \cong 0, \quad (h_{12}||h_{13}) \cong 0, \quad (h_{12}||h_{14}) \cong 0, \tag{2.65}$$

and also

$$\begin{aligned}
(h_{11}||h_{11}) \cong (h_{13}||h_{13}) &\cong \frac{1}{2}A, \\
(h_{12}||h_{12}) \cong (h_{14}||h_{14}) &\cong \frac{1}{2}B, \\
(h_{11}||h_{12}) \cong (h_{13}||h_{14}) &\cong \frac{1}{2}C,
\end{aligned} \tag{2.66}$$

where $A := (a||a)$, $B := (b||b)$, $C := (a||b)$. With these approximations the matrix $\mathcal{M}$ is given by

$$\mathcal{M} = \begin{pmatrix} C & \mathcal{O} \\ \mathcal{O} & C \end{pmatrix}, \tag{2.67}$$

where $\mathcal{O}$ is a zero 2×2 matrix and C equals

$$C = \frac{1}{2}\begin{pmatrix} A & C \\ C & B \end{pmatrix}. \tag{2.68}$$

Thus, $\mathcal{M}$ splits into two identical 2×2 matrices. Assuming $a \neq b$, $A \neq 0$, and $B \neq 0$, the explicit expressions for maximum likelihood estimators $\hat{A}_{1i}$ of the amplitudes A_{1i} are given by

$$\begin{aligned}
\hat{A}_{11} &= 2\frac{B(x||h_{11}) - C(x||h_{12})}{D}, \\
\hat{A}_{12} &= 2\frac{A(x||h_{12}) - C(x||h_{11})}{D}, \\
\hat{A}_{13} &= 2\frac{B(x||h_{13}) - C(x||h_{14})}{D}, \\
\hat{A}_{14} &= 2\frac{A(x||h_{14}) - C(x||h_{13})}{D},
\end{aligned} \tag{2.69}$$

where $D = AB - C^2$. The second partial derivatives of the log likelihood function with respect to A_{1i} are

$$\frac{\partial^2 \ln \Lambda_1'}{\partial A_{1i} \partial A_{1j}} = -\mathcal{M}_{ij}. \tag{2.70}$$

Since $a \neq b$, it follows from the Schwarz inequality that $D > 0$. Thus as $A > 0$ and $B > 0$ the matrix $\mathcal{M}$ is positive-definite. Therefore the extrema of the log likelihood function with respect to A_{1i} are the local maxima. The above maximum likelihood estimators of the amplitudes A_{1i} are substituted for the amplitudes A_{1i} in the likelihood function (2.61) giving the reduced normalized likelihood function $\Lambda_1'' = \exp(\mathcal{F}_1)$ where $\mathcal{F}_1$ is given by

$$\mathcal{F}_1 = \frac{B(x||h_{11})^2 + A(x||h_{12})^2 - 2C(x||h_{11})(x||h_{12})}{D} \tag{2.71}$$

$$+ \frac{B(x||h_{13})^2 + A(x||h_{14})^2 - 2C(x||h_{13})(x||h_{14})}{D}. \tag{2.72}$$

To obtain the maximum likelihood estimators of the signal parameters, first find the maximum of the functional $\mathcal{F}_1$ with respect to the frequency, the spin-down parameters, and the angles α and δ, then calculate the estimators of the amplitudes A_{1i} from the analytic formulae (2.69) with the correlations $(x||h_{1i})$ evaluated at the parameter values obtained by maximizing the functional $\mathcal{F}_1$. Therefore, filtering for the GW signal from a neutron star requires four linear filters. The same procedure applies to the second component of the signal. The formulae for the estimators of the amplitudes A_{2i} and the normalized reduced statistics $\mathcal{F}_2$ are obtained from the above formulae by replacing h_{1i} with h_{2i}. To consider the optimal detection of the whole two-component signal, remember that the eight amplitudes A_{1i} are not independent. They depend on five parameters: h_0, θ, ψ, ι, and Φ_0. To find the maximum likelihood estimators of the independent five parameters, one would need to maximize the total likelihood function (given by Eq. (2.57)) with respect to these parameters. This, however, leads to an intractable set of nonlinear algebraic equations which would need to be solved numerically, thereby increasing the computational cost of the signal search. Instead, an alternative is to form the statistic

$$\mathcal{F} = \frac{T_0}{S_h(f_0)}\mathcal{F}_1 + \frac{T_0}{S_h(2f_0)}\mathcal{F}_2. \tag{2.73}$$

This is the reduced likelihood function assuming the eight amplitudes are independent. First, maximize the functional $\mathcal{F}$ with respect to the frequency, spindown parameters, and angles α and δ, and calculate the eight amplitudes from the analytic formulae. Then, find the estimators of the five independent parameters from the estimators of the amplitudes using the least-squares method. The inverse of the Fisher matrix is used for the covariance matrix in the least-squares method. To announce the detection of the signal, the functional $\mathcal{F}$ must exceed a certain threshold calculated based on the false alarm probability that can be afforded. Once $\mathcal{F}$ is above the threshold, its magnitude determines the probability of signal detection. Consequently, the probability density function of $\mathcal{F}$ must be determined both when the signal is absent and present. More detail on the probability distribution functions can be found in section 2.4.1.3.

2.2.3.2 Data Analysis

Calculating the optimum statistics as efficiently as possible is crucial. One effective method is to utilize the speed of the *fast Fourier transform* (FFT). Consider the normalized reduced functional $\mathcal{F}_1$. The phase Φ of the signal can be expressed as (cf. Eq. (2.34))

$$\Phi(t) = 2\pi f_0[t + \Phi_m(t;\alpha,\delta)] + \Phi_s(t; f_0^{(k)}, \alpha, \delta), \tag{2.74}$$

where functions Φ_m and Φ_s are independent of the frequency parameter f_0. Define the following two integrals

$$\mathcal{F}_{1a} = \int_{-T_0/2}^{T_0/2} x(t)a(t)\exp[-i\Phi_s(t)]\exp\{-i2\pi f_0[t+\Phi_m(t)]\}dt, \tag{2.75}$$

$$\mathcal{F}_{1b} = \int_{-T_0/2}^{T_0/2} x(t)b(t)\exp[-i\Phi_s(t)]\exp\{-i2\pi f_0[t+\Phi_m(t)]\}dt. \tag{2.76}$$

The statistics $\mathcal{F}_1$ can be written in terms of the above integrals as

$$\mathcal{F}_1 = \frac{4}{T_o^2}\frac{B|F_{1a}|^2 + A|F_{1b}|^2 - 2C\Re(F_{1a}^* F_{1b})}{D}. \tag{2.77}$$

Introduce a new time coordinate t_b

$$t_b(t) = t + \Phi_m(t). \tag{2.78}$$

From the explicit expression for the phase Φ given by Eq. (2.34), the time shift Φ_m and its time derivative $\dot{\Phi}_m$ can be estimated as follows

$$\begin{aligned} |\Phi_m(t)| &\lesssim \frac{R_{\rm ES}}{c} \simeq 5\times 10^2 \text{ s}, \\ |\dot{\Phi}_m(t)| &\lesssim \frac{\Omega_0 R_{\rm ES}}{c} \simeq 1\times 10^{-4}. \end{aligned} \tag{2.79}$$

Assuming a maximum observation time $T_0 = 120$ days, it is approximated as

$$T_b := t_b(T_o) \cong T_o, \quad \frac{dt}{dt_b} \cong 1. \tag{2.80}$$

Thus, in the new time coordinate, the integrals F_{1a} and F_{1b} can be approximated by

$$F_{1a} \cong \int_{-T_o/2}^{T_o/2} x[t(t_b)]a[t(t_b)]\exp[-i\Phi_s(t[t_b])]\exp(-i2\pi f_0 t_b)dt_b, \tag{2.81}$$

$$F_{1b} \cong \int_{-T_o/2}^{T_o/2} x[t(t_b)]b[t(t_b)]\exp[-i\Phi_s(t[t_b])]\exp(-i2\pi f_0 t_b)dt_b. \tag{2.82}$$

With the new time coordinate, the two integrals (2.81) and (2.82) are Fourier transforms of the functions $x[t(t_b)]a[t(t_b)]\exp\{-i\Phi_s[t(t_b)]\}$ and $x[t(t_b)]b[t(t_b)]\exp\{-i\Phi_s[t(t_b)]\}$, respectively. To calculate these integrals for a given set of phase parameters, perform the following numerical operations. For the chosen values of the parameters α and δ, resample the original time series according to the formula (2.78) and then multiply the resampled time series $x[t_b]$ by functions $a(t_b)\exp[-i\Phi_s(t[t_b])]$ and $b(t_b)\exp[-i\Phi_s(t[t_b])]$. Then calculate the two Fourier transforms using the FFT algorithm. This resampling technique, proposed by Schutz [114], is considered an effective data analysis tool for GW searches from periodic sources [51]. Alternatively, define new spin-down parameters

$$f_k := \frac{f_0^{(k)}}{f_0}, \quad k = 1,\ldots,s, \tag{2.83}$$

and introduce a different time coordinate

$$t'_b(t) = t + \Phi_m(t) + \Phi'_s(t), \tag{2.84}$$

where

$$\Phi'_s(t) = \sum_{k=1}^{s} f_k \frac{t^{k+1}}{(k+1)!} \tag{2.85}$$

and perform the resampling process according to the formula (2.84). The functions a, b, and consequently A, B, C, and D are known and depend on δ, α, and the observation time. Their values can be calculated and stored for a fine grid of neutron star positions on the sky and appropriate observation times before the data analysis is conducted. The normalized reduced functional for the second component of the signal can be calculated similarly. Here the corresponding Fourier transforms are given by

$$F_{2a} \cong \int_{-T_o/2}^{T_o/2} x[t(t_b)]a[t(t_b)] \exp[-i2\Phi_s(t[t_b])] \exp(-i4\pi f_0 t_b) dt_b, \tag{2.86}$$

$$F_{2b} \cong \int_{-T_o/2}^{T_o/2} x[t(t_b)]b[t(t_b)] \exp[-i2\Phi_s(t[t_b])] \exp(-i4\pi f_0 t_b) dt_b. \tag{2.87}$$

The statistics $\mathcal{F}$ for the whole signal is then calculated from the formula

$$\mathcal{F} = \frac{4}{S_h(f_0)T_o} \frac{B|F_{1a}|^2 + A|F_{1b}|^2 - 2C\Re(F_{1a}F^*_{1b})}{D} \tag{2.88}$$

$$+ \frac{4}{S_h(2f_0)T_o} \frac{B|F_{2a}|^2 + A|F_{2b}|^2 - 2C\Re(F_{2a}F^*_{2b})}{D}. \tag{2.89}$$

The statistics $\mathcal{F}$ needs to be calculated on a multidimensional grid of parameter values covering sufficiently densely the parameter space, and compared against a threshold.

2.2.3.3 The $\mathcal{F}$- statistic

It has been demonstrated [115] that the quantity $2\mathcal{F}$ follows a probability distribution of a χ^2 with four degrees of freedom in the absence of a signal and that of a non-central χ^2 with a signal present with a non-centrality parameter

$$\lambda := d^2 = (h|h) \tag{2.90}$$

where d is proportional to the signal amplitude. The probability distributions $p_{\text{noise}}(2\mathcal{F})$ and $p_{\text{signal+noise}}(2\mathcal{F}; d)$ are hence

$$p_{\text{noise}}(2\mathcal{F}) = \frac{1}{4}(2\mathcal{F})^{-1/2} e^{-(2\mathcal{F})/2}, \tag{2.91}$$

$$p_{\text{signal+noise}}(2\mathcal{F}; d) = \left(\frac{2\mathcal{F}}{d}\right)^{1/2} I_1\left(d\sqrt{2\mathcal{F}}\right) e^{-\frac{1}{2}(2\mathcal{F}) - \frac{1}{2}d^2}, \tag{2.92}$$

where I_1 is a modified Bessel function of the first kind. As discussed in section 2.2.3.2, $2\mathcal{F}$ can be used as a detection statistic, where a threshold $2\mathcal{F}_0$ can be chosen to satisfy a desired false alarm probability

$$\begin{aligned}\mathrm{CDF}_{\mathrm{noise}}[2\mathcal{F}_0] &= \int_0^{2\mathcal{F}_0} p_{\mathrm{noise}}(2\mathcal{F})\, d(2\mathcal{F}) = 1-\alpha, \\ &= 1 - \left(1 + 2\mathcal{F}_0 + \frac{1}{2}(2\mathcal{F}_0)^2 + \frac{1}{6}(2\mathcal{F}_0)^3\right) e^{-2\mathcal{F}_0}, \end{aligned} \tag{2.93}$$

and where the probability of detection for a given d is

$$P_{\mathrm{detection}}(d, 2\mathcal{F}_0) = \int_{2\mathcal{F}_0}^{\infty} p_{\mathrm{signal+noise}}(2\mathcal{F}; d)\, d(2\mathcal{F}). \tag{2.94}$$

The formalism above describes a time-domain implementation, but a narrow-band frequency implementation [116] has been used extensively in LIGO searches. In the case of known pulsars for which optical or X-ray observations of pulsar wind nebulae allow inference of ι and ψ, an improved version of the $\mathcal{F}$-statistic known as the $\mathcal{G}$-statistic can be applied to get a slight gain in sensitivity [117]. Although originally derived in a frequentist, log-likelihood approach, the $\mathcal{F}$-statistic can also be derived in a Bayesian framework [118] by setting an unphysical prior (non-isotropic in stellar orientation).

2.3 Continuous Wave searches

As detailed in Sec. 2.1, the most plausible sources of CW gravitational radiation detectable by ground-based detectors are rapidly spinning neutron stars within our galaxy. Search strategies for CW radiation are highly dependent on the *a priori* information about the source. CW searches can be categorized into three broad types: [119] (1) *targeted searches*, where both the star's position and rotation frequency are known, such as known radio, X-ray, or γ-ray pulsars; (2) *directed searches*, where the position of the star is known but the rotation frequency is not, for example, a non-pulsating X-ray source located at the center of a supernova remnant; and (3) *all-sky searches* aimed at discovering unknown neutron stars. The parameter space to be searched expands significantly across these categories. In each scenario, the star might be isolated or part of a binary system. Given the anticipated weak signal strengths, it is important to integrate data over extended observation periods to enhance the likelihood of signal detection. The amount of known information about the source determines the integration approach. Generally, the more detailed the knowledge, the more computationally manageable it is to perform coherent data integration (maintaining phase information) over extended observation times, as described below.

2.3.1 Performance measures

A significant challenge in CW searches lies in determining the parameter space of signals to be covered and selecting the appropriate detection statistic. These elements of CW search design must generally be balanced against each other. Given the constraints of computational resources, it is often necessary to

choose between a sensitive (but computationally intensive) detection statistic over a limited (and computationally less demanding) parameter space, or a broader (but more computationally costly) parameter space with a sub-optimal (and computationally cheaper) detection statistic.

2.3.1.1 Sensitivity depth

The CW search sensitivity depth [120, 121] is a scalar that functions similarly to a signal-to-noise ratio. As a measure of signal strength, an *upper limit at confidence* C is calculated for the GW amplitude of the signal, typically denoted as h_0. The upper limit h_0^C is determined using various methods [121, 122], which generally adhere to the following procedure

1. Establish a criterion for detecting CW signals using a given search algorithm. Common practice is to consider the maximum detection statistic $X^{\max}$ found by a search as the most promising detection candidate. Apply the algorithm to numerous synthesized data sets containing only noise, record the $X^{\max}$ from each data set, and set a threshold X^* such that only a fraction A of the $X^{\max}$ satisfy $X^{\max} > X^*$. Here, A represents the false alarm probability

   > The probability of falsely claiming a detection by finding $X^{\max} > X^*$, despite the data containing no signal. This probability is typically small; conventionally, CW searches set a false alarm threshold of $A = 1\%$.

2. With a fixed h_0, draw random sets of values for the CW signal model parameters. Amplitude parameters (other than h_0) are drawn from their natural priors; phase parameters are drawn from the search parameter space.

3. For each set of model parameter values, generate a CW signal time-series $h(t)$ in software using the given parameter values. Add to each $h(t)$ a realistic representation of the detector noise, either by using real detector data or by simulating stationary Gaussian noise with power spectral density S_h (assumed constant over the narrow bandwidth of CW signals).

4. Analyze each noisy $h(t)$ time series using the CW search algorithm. For each noisy $h(t)$, determine if the simulated CW signal would have been detected using the established criteria, i.e., whether the $X^{\max}$ from the search of each noisy $h(t)$ exceeds X^*. The fraction of CW signals considered detected is the *detection confidence* C. The h_0 fixed in step 2 is then interpreted as an upper limit h_0^C with C confidence; if the data contains a signal with amplitude h_0, there is confidence C of detecting it.

5. Adjust h_0 and repeat steps 2-4 until C converges to the desired confidence level, typically 90% or 95%. Published CW searches generally report upper limits h_0^C as a function of search frequency f as their primary scientific result.

The sensitivity depth takes the ratio of an upper limit, given by h_0^C, to an estimate of the noise in the detector, given by S_h. It is defined as

$$\mathcal{D} = \frac{\sqrt{S_h/\mathrm{Hz}}}{h_0^C}. \tag{2.95}$$

Here, S_h is defined as the single-sided power spectrum within the same frequency band as the CW signals used to determine h_0^C, and it is calculated by taking the harmonic mean of the data over time and across multiple detectors. This averaging over time provides a representative value of the detector's power spectrum, which can vary over short periods during an observation. Short-duration non-Gaussian "glitches" in the noise (lasting less than 1 second) generally do not impact long-duration CW searches unless they are intense and short enough (similar to a Dirac delta function) to affect a wide frequency band in their Fourier transform [123].

Since a smaller h_0^C signifies a more sensitive search, $\mathcal{D}$ increases with search sensitivity. By isolating the performance of the detector, specifically its noise power spectrum, $\mathcal{D}$ measures the contribution of the CW search algorithm to the overall search sensitivity. Typically, $\mathcal{D}$ stays roughly constant for a given CW search, excluding frequency bands where S_h is negatively impacted by excessive detector noise, such as instrumental line artifacts.

2.3.1.2 Parameter-space breadth

To complement $\mathcal{D}$ as a measure of CW search sensitivity, the concept of breadth $\mathcal{B}$ is introduced as a measure of CW parameter-space coverage

$$\mathcal{B} = \int_{\mathcal{P}} d^p\vec{p}\sqrt{g(\vec{p})}. \tag{2.96}$$

Here, $\mathcal{P}$ represents the CW search parameter space, from which parameter vectors $\vec{p} \in \mathcal{P}$ are drawn. The function $g(\vec{p})$ is the determinant of the *parameter-space metric* $g_{ij}(\vec{p})$ [124]. The metric $g_{ij}(\vec{p})$ provides a distance measure in the parameter space. Specifically, if a CW signal is present in the data with best-fit parameters $\vec{p}$, and an attempt is made to match that signal to a model waveform with parameters $\vec{p}+\Delta\vec{p}$, the mismatch $\mu = g_{ij}(\vec{p})\Delta p_i \Delta p_j$ quantifies the expected loss in signal power. If $\Delta\vec{p} = 0$, the signal is perfectly matched, resulting in the optimal maximum detection statistic. For $\Delta\vec{p} \neq 0$, the detection statistic is expected to decrease by a factor $\approx 1 - \mu$ for small μ [125].

The metric quantitatively measures the "closeness" of signals in parameter space in terms of recovering signal power, with its determinant quantifying the "density" of the parameter space. Thus, $\mathcal{B}$ represents the number of CW model waveforms (or *templates*) required to search a given parameter space thoroughly. However, $\mathcal{B}$ is not rescaled to provide actual template counts for a given CW search, as this is considered an implementation detail of the algorithm. Ideally, the algorithm would use the minimum number of templates necessary to cover the parameter space, ensuring the mismatch remains under a pre-established maximum [126]. It has been shown [127] that for observing time-spans longer than a day, the CW parameter-space metric is largely independent of the amplitude parameters of the CW signal. Consequently, only the CW phase is needed to compute the metric; indeed, it suffices to use a simplified, approximate phase. Using the *phase metric* approximation [127], the metric is computed from:

$$g_{ij}(\vec{p}) = \frac{1}{T}\left[\int_T dt \frac{\partial\phi(t)}{\partial p_i}\frac{\partial\phi(t)}{\partial p_j}\right]_{\vec{p}} - \frac{1}{T^2}\left[\int_T dt \frac{\partial\phi(t)}{\partial p_i}\right]_{\vec{p}}\left[\int_T dt \frac{\partial\phi(t)}{\partial p_j}\right]_{\vec{p}}, \tag{58}$$

where the integration spans the observation period T. For the parameter-space breadth computation, the parameter-space metric of a *fully-coherent* CW signal is assumed, where $\phi(t_{\rm det})$ matches the data across the entire observation. Not all detection statistics meet this criterion; often, a trade-off is made by sub-optimal detection statistics to relax this restriction. The purpose of $\mathcal{B}$ is to quantify the size of the CW signal model's parameter space, while the CW search size (in terms of the number of templates) is considered an implementation detail of the search. CW data analysts must often balance sensitivity and parameter-space coverage in their CW search design, with the trade-off reflected in the $\mathcal{D}$ and $\mathcal{B}$ achieved for that search. An optimal choice could be to use a detection statistic that balances sensitivity (increasing depth) with computational efficiency, allowing a wider parameter space to be searched (increasing breadth).

Frequency and spindown parameter space

The phase metric over the frequency and spindown parameters is given by [128]

$$g_{f^{(r)},f^{(s)}} = \frac{4\pi^2(r+1)(s+1)}{(r+2)!(s+2)!(r+s+3)!}T^{r+s+2}, \tag{2.97}$$

where $f^{(r)}, f^{(s)}$ are the rth, sth spindown parameters respectively. Note that $g_{f^{(r)},f^{(s)}}$ is independent of $f^{(s)}$. The square root of the determinants of this metric, up to the 2nd spindown, are

$$\mathcal{B}_f = \sqrt{g(f)} = \frac{\pi T}{\sqrt{3}}, \tag{2.98}$$

$$\sqrt{g(f,\dot{f})} = \frac{\pi^2 T^3}{6\sqrt{15}}, \tag{2.99}$$

$$\sqrt{g(f,\dot{f},\ddot{f})} = \frac{\pi^3 T^6}{360\sqrt{105}}. \tag{2.100}$$

Equation (2.97) assumes the observation runs over $t_{\rm det} \in [0, T]$, while the subsequent equations are independent of this choice. Equation (2.98) provides the raw frequency breadth $\mathcal{B}_f$. To separate the contributions from each spin-down parameter, their (raw) breadths are defined as ratios to the breadth of the preceding spin-down

$$\mathcal{B}_{\dot{f}} = \frac{1}{\mathcal{B}_f}\int df\, d\dot{f}\sqrt{g(f,\dot{f})} = \frac{\pi T^2}{6\sqrt{5}}\left|\dot{f}\right], \tag{2.101}$$

$$\mathcal{B}_{\ddot{f}} = \frac{1}{\mathcal{B}_{\dot{f}}}\int df\, d\dot{f}\, d\ddot{f}\sqrt{g(f,\dot{f},\ddot{f})} = \frac{\pi T^3}{60\sqrt{7}}\left|\ddot{f}\right]. \tag{2.102}$$

Using the notation of [127], for a parameter p with parameter space $[p_0, p_1] \cup [p_2, p_3] \cup \ldots$, define:

$$|p^q] \equiv p_1^q - p_0^q + p_3^q - p_2^q + \ldots. \tag{2.103}$$

The breadths in $f^{(s)}$ scale as T^{s+1}, consistent with the typical spacings $\Delta f^{(s)} \propto T^{-s-1}$ used to construct a rectangular search grid in these parameters.

Sky parameter space

The metric over the sky arises from the Solar System Rømer delay $\Delta R_0(t_{\rm det})$ [117]. The detector position vector $\vec{r}(t_{\rm det})$ can be approximated with a Ptolemaic-like orbit (cf. [129]), assuming the Earth's orbit is

circular and co-planar with its equator, and ignoring the relative phase differences between the Earth's sidereal and orbital motions, which should be immaterial over long observation periods. With these assumptions, the sky component of the CW phase simplifies to

$$\phi_{\rm sky}(t_{\rm det}) = 2\pi f \cos\delta \left[\tau_s \cos(\alpha - \Omega_s t_{\rm det}) + \tau_0 \cos(\alpha - \Omega_0 t_{\rm det})\right]. \tag{2.104}$$

Here, $\tau_s \approx 2.13 \times 10^{-2}$ lt-s is the Earth's radius, $\tau_0 \approx 4.99 \times 10^2$ lt-s is the orbital radius of the Earth around the Sun, and $\Omega_s \approx 7.27 \times 10^{-5}$ Hz, $\Omega_0 \approx 1.99 \times 10^{-7}$ Hz are the respective sidereal and orbital angular frequencies. Combining Eqs. (2.96), (58), and (2.104), and noting that $\tau_s \ll \tau_0$ and $\Omega_0 \ll \Omega_s$, the expression for $\mathcal{B}_{\rm sky}$ is

$$\mathcal{B}_{\rm sky} = \frac{8}{3}\pi^3 f^2 \tau_0^2 \Bigg\{ 1 - {\rm sinc}^2(\Omega_0 T) - 2{\rm sinc}^2\left(\frac{\Omega_0 T}{2}\right) \tag{2.105}$$

$$+ 2{\rm sinc}(\Omega_0 T)\ {\rm sinc}^2\left(\frac{\Omega_0 T}{2}\right) + \frac{4\tau_s}{\tau_0}{\rm sinc}\left(\frac{\Omega_s T}{2}\right) + \mathcal{O}\left[\left(\frac{\tau_s}{\tau_0}\right)^2\right]\Bigg\}^{1/2}, \tag{2.106}$$

where ${\rm sinc}(x) = \sin(x)/x$. Note that $\mathcal{B}_{\rm sky}$ converges to $8\pi^3 f^2 \tau_0^2/3$ for $\Omega_0 T \gg 1$, i.e., once the observation time spans many years. For CW searches covering patches of the sky, $\mathcal{B}_{\rm sky}$ is rescaled by the fraction of the sky covered, assuming the template density is isotropic over the sky for long observations. Given several types of raw breadths defined previously, the overall breadth is found by multiplying these factors and integrating over f

$$\mathcal{B} = \int df \prod \mathcal{B}_{\ldots}, \tag{2.107}$$

where the product is taken over those $\mathcal{B}_{\ldots}$ relevant to a particular search.

2.4 Search algorithms

Given a model for the CW signal waveform (Sec. 2.2), the optimal strategy is to match filter the entire data set against the model over the parameter space of interest. This method is feasible when all phase parameters of the signal are well known, such as in targeted and narrow-band searches for known pulsars. However, it becomes impractical over a wide parameter space. To illustrate, consider an all-sky search of one year of data, covering $f \in [50, 1000]$ Hz and $\dot{f} \in [-10^{-8}, 0]$ Hz s^{-1}. Using the parameter-space breadth formulae from Sec. 4.2, we get $\mathcal{B}_{\rm sky} = 9.9 \times 10^{12}$, $\mathcal{B}_f = 4.0 \times 10^{10}$, $\mathcal{B}_{\dot{f}} = 2.3 \times 10^6$, and $\mathcal{B} = 9.2 \times 10^{29}$. Assuming $\mathcal{B}$ approximates the number of matched filters needed, the computational cost on current hardware is $\approx 2.3 \times 10^{19}$ years, which is clearly impractical. To address this, CW search algorithms use a *hierarchical structure* [51]. Initially, the entire data set (time-span T) is divided into N *segments*, each spanning a coherence time $T_{\rm coh} \ll T$. In the first *coherent stage*, matched filters are applied to each segment individually. The number of filters required generally scales with $T_{\rm coh}^{\delta}$ where $\delta \gtrsim 6$ [130]. This steep scaling is computationally expensive for a fully-coherent search, but choosing a manageable $T_{\rm coh}$ makes the computational cost feasible.

In the second *semi-coherent stage*, an algorithm combines the matched filter results from the N segments. This algorithm sums the N matched filter results over the same phase parameter space, allowing phase jumps between segments while requiring consistent frequency. The phase can jump by a random offset from segment to segment, while the frequency remains consistent across segments

$$\phi(t)_{N+1} = \phi(t)_N + \phi_{\text{random}}, \tag{2.108}$$

$$\begin{aligned} 2\pi f(t)_{N+1} &= \frac{d\phi(t)_N}{dt} + \frac{d\phi_{\text{random}}}{dt} \\ &= \frac{d\phi(t)_N}{dt} + 0 = 2\pi f(t)_N, \end{aligned} \tag{2.109}$$

where $\cdot_{N+1}$ and $\cdot_N$ label the $(N+1)$th and Nth segments, respectively. Relaxing phase consistency reduces sensitivity and makes the detection statistic more susceptible to instrumental line artifacts [131]. However, the computational cost of the semi-coherent stage scales as $N\eta T_{\text{coh}}^{\delta}$ with $\eta \ll \delta \approx \delta$ [130], saving a factor of $\sim N^{\delta-\eta} T_{\text{coh}}^{-\delta}$ in computational cost compared to a fully-coherent analysis. This significant saving often makes hierarchical searches of year-long data sets feasible. Given a fixed computing budget [130], a hierarchical search is often more sensitive than a fully-coherent analysis, as it can analyze all available data efficiently, whereas a fully-coherent search would be limited to a subset of the data, thereby reducing its sensitivity.

2.4.1 Coherent algorithms

Figure 2.3 illustrates $\mathcal{D}$ and $\mathcal{B}$ for each search as a function of the coherence time T_{coh} used in the initial coherence stage. If a search employs multiple coherence times, the maximum coherence time $T_{\text{coh,max}} = \max T_{\text{coh}}$ is displayed. It is evident that search sensitivity improves with longer coherence times, while parameter-space coverage diminishes due to the sharp rise in computational cost associated with increasing T_{coh}. The four coherent algorithms identified in the figure are discussed in the following section.

2.4.1.1 Frequency-domain power

Over a "short enough" time T_{coh}, the CW signal frequency remains nearly constant. The simplest coherent algorithm for this scenario involves computing the discrete Fourier transform of a data segment and then determining the power of the bin where the CW signal is expected. The criterion for T_{coh} to be "short enough" so that the CW signal power is concentrated in one bin is given by

$$T_{\text{coh}}^2 \leq \frac{6\sqrt{5\mu T_{\text{coh}}}}{\pi \max\{f\} \max\{\tau_s, a_p\} \max\{\Omega_s^2, \Omega_p^2\}} \tag{2.110}$$

where μT_{coh} sets the fraction of signal power allowed to be lost for a specific T_{coh}, $\max\{f\}$ is the maximum frequency over the parameter space, $\max\{\tau_s, a_p\}$ is the maximum of the Earth's radius and the searched binary orbit's projected semi-major axis (if any); and $\max\{\Omega_s^2, \Omega_p^2\}$ is the maximum of the square of either the Earth's sidereal angular frequency or (if any) the searched angular frequency of a binary orbit.

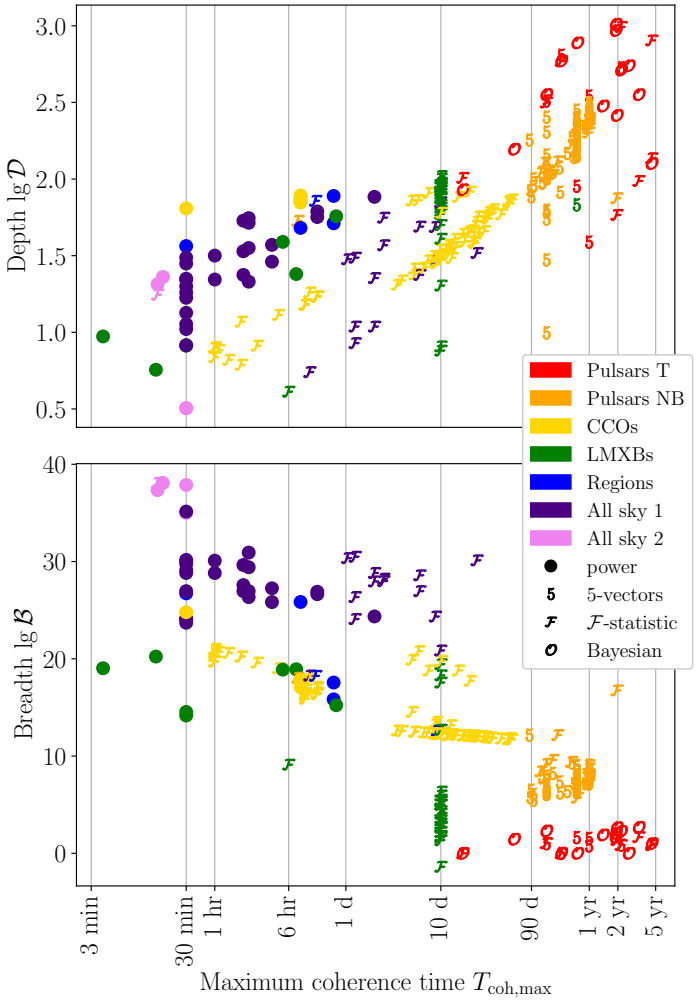


FIGURE 2.3: Sensitivity depth $\mathcal{D}$ (top panel) and parameter-space breadth $\mathcal{B}$ (bottom panel) of 297 CW searches, plotted against the maximum coherence time $T_{\mathrm{coh,max}}$ used in each search. Markers denote the coherent algorithm employed by each search. Taken from [132].

GW data, divided into segments of time-span T_{coh} as described above and Fourier transformed, serves as a common input data product not only for semi-coherent algorithms that sum power in each segment but also for algorithms such as the $\mathcal{F}$-statistic (Sec. 2.2.3). Common file formats include the Short Fourier Transform (SFT) [133], and the Short FFT DataBase (SFDB) [134].

2.4.1.2 5-vectors algorithm

The 5-vectors method [135, 136] exploits the following property of the CW signal: when the phase parameters of the signal are fully specified (i.e., the functions $h_+(t)$ and $h_\times(t)$ in Eq. (1.43) are precisely known), only the modulation from the detector responses $F_+(t)$ and $F_\times(t)$ remains. The strain amplitude can be expressed as

$$h(t) = h_0 \boldsymbol{A} \cdot \boldsymbol{W} e^{i(\Omega_s t + \phi_0)} \tag{2.111}$$

where Ω_s is the signal frequency in the SSB frame, $\boldsymbol{A}$ is decomposable into plus- and cross-polarized terms that depend on complex amplitudes H_+ and $H_\times$

$$\boldsymbol{A} = H_+ \boldsymbol{A}^+ + H_\times \boldsymbol{A}^\times \tag{2.112}$$

Here, $\boldsymbol{A}^+$ and $\boldsymbol{A}^\times$ are expressed using trigonometric functions [32]. The vector $\boldsymbol{W}$ is a five-component set of basis functions, indexed by $k = -2, -1, 0, 1, 2$:

$$\boldsymbol{W}_k = e^{-ik\Theta} \tag{2.113}$$

where Θ is the detector's local sidereal time in radians. The data stream $x(t)$ can also be decomposed using these basis functions

$$\boldsymbol{X} = \int_T x(t) \boldsymbol{W} e^{-i\Omega_s t} dt \tag{2.114}$$

A detection statistic is then constructed using a weighted sum of the squared projections

$$S = c_+ |\hat{h}_+|^2 + c_\times |\hat{h}_\times|^2 \tag{2.115}$$

where the projections are defined by

$$\hat{h}_+ = \frac{\boldsymbol{X} \cdot \boldsymbol{A}^+}{|\boldsymbol{A}^+|^2}, \quad \hat{h}_\times = \frac{\boldsymbol{X} \cdot \boldsymbol{A}^\times}{|\boldsymbol{A}^\times|^2} \tag{2.116}$$

Empirical studies [32] indicate that the best performance for known ι and Ψ is achieved with weightings $c_{+,\times} = |\boldsymbol{A}^{+,\times}|^4$, and the signal amplitude estimation can be obtained from

$$\hat{h}_0 = \sqrt{|\hat{h}_+|^2 + |\hat{h}_\times|^2} \tag{2.117}$$

To date, the 5-vectors method has primarily been used as a fully-coherent algorithm for targeted and narrow-band searches for known pulsars. The phase demodulation of the signal using the known $h_+(t)$ and $h_\times(t)$ is efficiently handled using the Band-Sampled Data (BSD) [137] framework, which provides band-limited data, heterodyned at regular intervals (typically 10 Hz), and down-sampled to speed up subsequent computations.

2.4.1.3 Coherent $\mathcal{F}$-statistic

As discussed in section 2.2.3, the $\mathcal{F}$-statistic is the log-likelihood function of the observed data given a CW phase modulation, maximised over the four amplitudes of Eq. (2.43). This statistic has been employed in various CW searches: for known pulsars (as a fully-coherent search), directional targets such as CCOs and LMXBs (both fully-coherently and as the first stage of a hierarchical pipeline), and all-sky searches.

The $\mathcal{F}$-statistic has specific statistical properties. In the absence of a signal and assuming Gaussian noise, the value of $2\mathcal{F}$ follows a central χ^2 distribution with four degrees of freedom. When a signal is present,

$2\mathcal{F}$ follows a non-central χ^2 distribution with four degrees of freedom and a non-centrality parameter ρ^2. Here, ρ^2 is the optimal signal-to-noise ratio when the signal and template are perfectly matched [117]. Due to these properties, explicit values of the $\mathcal{F}$-statistic quoted in the literature are usually presented as $2\mathcal{F}$.

Multiple software implementations exist to compute the $\mathcal{F}$-statistic. One time-domain implementation [138] uses the "resampling" technique introduced in [117]. This method employs the Fast Fourier Transform to compute N values of $2\mathcal{F}$ at regularly spaced frequencies simultaneously. The computational cost of resampling scales as $\mathcal{O}(\log N)$, compared to $\mathcal{O}(N)$ for computing each $2\mathcal{F}$ value individually. The LALSuite [139] software package includes an independent implementation of the resampling technique [140]. LALSuite also features another frequency-domain technique known as "demodulation," which efficiently computes individual $2F$ values [141, 142].

For neutron star sources in binary orbits, the signal phase must be de-modulated according to $\Delta_R(t_{NS})$ (Sec. 2.2.2). An alternative method [143] involves computing the $\mathcal{F}$-statistic without including $\Delta_R(t_{NS})$ in the phase; instead, the signal is divided into multiple side-bands using the Jacobi-Anger expansion [cf. [144]]. A subset of these side-bands can then be summed to reconstruct the $\mathcal{F}$-statistic.

2.4.1.4 Bayesian inference

Bayes' theorem provides the posterior probability of a model given observed data, derived from the prior probability for the model, the likelihood of the data given the model, and the evidence (or marginal likelihood). If the prior probability distributions align with the expected population of signals, Bayesian inference offers the most robust detection statistic [145]. The $\mathcal{F}$-statistic, for example, does not meet this criterion; the maximization over the A^i amplitudes implicitly assumes unrealistic prior distributions for the underlying model parameters h_0 and $\cos\iota$ [146].

A Bayesian inference pipeline is employed for surveys of known pulsars. Unlike ad-hoc constructed statistics, Bayesian inference's flexibility allows for the use of the intricate pulsar timing model [147], including eccentric binary orbits and irregular timing noise [148]. Initially, the pipeline heterodynes the data at the specified CW phase inferred from the pulsar ephemeris, then computes posterior probabilities on the four physical amplitude parameters $h_0, \cos\iota, \psi, \phi_0$. The first implementation of this pipeline [149] utilized a Markov Chain Monte Carlo (MCMC) to compute the posterior probability. A more efficient version [150] employs nested sampling, enabling searches over small ranges of phase parameters, quantifying detection significance, and performing model comparison. It also supports searches at both harmonics of the CW signal [151] and models GW polarizations predicted by non-GR theories of gravity [152]. This implementation is integrated into LALSuite [139]. A third generation of the pipeline [153], written in Python, can utilize various Bayesian inference solvers via [154].

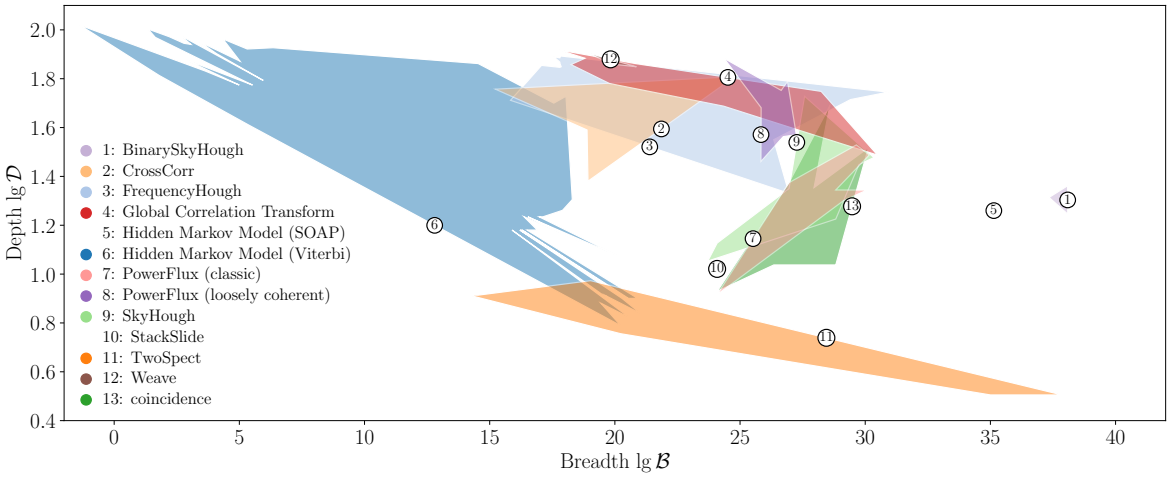


FIGURE 2.4: Sensitivity depth $\log \mathcal{D}$ versus parameter-space breadth $\log \mathcal{B}$ for 297 CW searches. Plotted are polygons, each of whose vertices indicates a CW search using a given semi-coherent algorithm. Names of all semi-coherent algorithms are listed in the legend. Each polygon is given a numeric label, placed on an arbitrary edge, corresponding to the legend. Where only one CW search uses a particular semi-coherent algorithm, no polygon is plotted and instead the numeric label marks the search $(\log \mathcal{B}, \log \mathcal{D})$. Taken from [132].

2.4.2 Semi-coherent algorithms

The primary distinction between a coherent framework and a semi-coherent one lies in their approach to data utilization. A coherent framework uses the entire data time-span T_{coh} in a single, unified analysis. In contrast, a semi-coherent framework divides the data into multiple segments, each of duration T_{semicoh}. Each segment is individually processed using a specific algorithm, and the results from all segments are subsequently combined to form a single detection statistic. Figure 2.4 plots $\mathcal{D}$ versus $\mathcal{B}$. In this figure, the 13 semi-coherent algorithms used by the searches are highlighted by plotting polygons whose vertices represent the searches. These polygons illustrate the area in $\mathcal{B}$–$\mathcal{D}$ space that each algorithm is demonstrably capable of operating over. This provides insight into the typical configurations and their flexibility for each algorithm. For instance, the Viterbi Hidden Markov Model (HMM) algorithm (Sec. 2.4.2.7) can be configured in various ways due to its computational efficiency, whereas the TwoSpect algorithm (Sec. 2.4.2.6), specialized for neutron stars in binary systems, is adaptable for both all-sky surveys (higher $\mathcal{B}$) and directed targets (lower $\mathcal{B}$). Many algorithms traditionally focus on specific parameter spaces. The following algorithms are typically used for all-sky searches for isolated neutron stars: coincidence (Sec. 2.4.2.1), SkyHough and FrequencyHough (Sec. 2.4.2.2), StackSlide, Global Correlation Transform (Sec. 2.4.2.3), and PowerFlux (and its loosely coherent extension; Sec. 2.4.2.4). CrossCorr (Sec. 2.4.2.5) focuses on LMXBs, particularly Sco X-1, while BinarySkyHough (Sec. 2.4.2.2) is designed for all-sky searches for neutron stars in binaries. Relatively newer algorithms, such as Weave (Sec. 2.4.2.3) and the SOAP HMM algorithm (Sec. 2.4.2.7), have only been used in a few searches to date. Figure 2.5 illustrates the maximum coherence time $T_{\text{coh,max}}$ usually chosen for each semi-coherent algorithm. (Note that $T_{\text{coh,max}}$ represents the longest coherence time used by a search, in cases where multiple coherence lengths are employed.) Seven of the 13 semi-coherent algorithms use power (Sec. 2.4.1.1) as the first-stage coherent algorithm; four utilize the $\mathcal{F}$-statistic (Sec. 2.4.1.3); and two may use

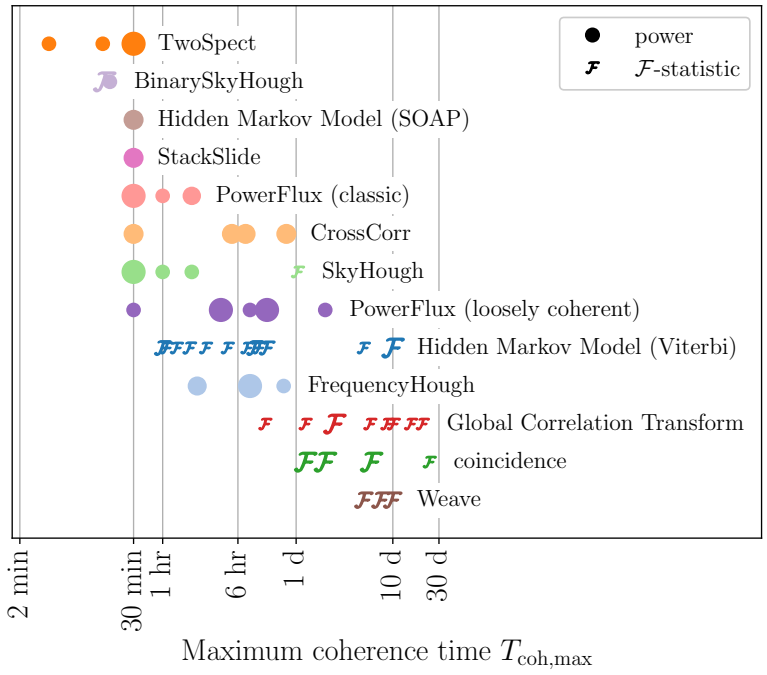


FIGURE 2.5: Maximum coherence times $T_{\mathrm{coh,max}}$ chosen by the semi- coherent algorithms used in 297 CW searches. Colours are arbitrary but are the same as in Fig. 2.4; markers denote the coherent algorithm used by the hierarchical pipelines; marker size indicates the relative popularity of a particular choice of $T_{\mathrm{coh,max}}$ for a given algorithm. Taken from [132].

either method. For semi-coherent algorithms using power, a coherence time of $T_{\mathrm{coh}} = 30$ minutes is a popular choice because it satisfies Eq. 2.110 up to $f \lesssim 2\,\mathrm{kHz}$ [155], making it the standard choice for the time-span of SFTs. Longer coherence times can be used when summing power at lower frequencies; however, searches over binary orbit parameters may require shorter SFTs if $\Omega_P < \Omega_S$ (Eq. 2.110). Semi-coherent algorithms that use the $\mathcal{F}$-statistic are generally not constrained by Eq. 2.110, though the computational cost increases significantly with T_{coh}.

The remainder of this section outlines the semi-coherent algorithms referenced in Figs. 2.4 and 2.5. Each algorithm essentially aims to demodulate the CW signal frequency accurately while allowing some flexibility in the CW signal phase, as described in Eq. 2.109. Despite this common goal, the algorithms differ significantly in their conceptual foundations and design choices.

2.4.2.1 Coincidence-based algorithms

The simplest method for combining the N segments is to avoid doing so; coincidence-based algorithms do not attempt to sum results across segments from the coherent stage. Instead, each segment is treated as an independent search, and coincidences are identified among significant candidates from each segment [138]. Candidates may be considered coincident by binning their parameters over a grid, thus grouping candidates with similar parameters [156].

The coincidence method may be less sensitive than other semi-coherent methods. For example, a weaker signal might become detectable by accumulating signal power over segments, but it could be missed

if it does not pass a per-segment threshold. However, this method has the advantage of avoiding the computational expense of the second semi-coherent stage. Additionally, it can be applied to data in (relatively) real time—as soon as T_{coh} worth of data is collected and prepared—rather than having to wait until the end of the run for all data to be collected. The coincidence method has been used in all-sky searches for isolated pulsars.

2.4.2.2 Hough transform algorithms

The Hough transform [157] is a pattern recognition algorithm originally developed to detect particle tracks in bubble chamber photographs. Consider analyzing a two-dimensional image (x, y) to identify straight line patterns of the form $y = mx + c$. The Hough transform maps a point (x, y) to the set $\{(m, c)\}$ of all lines passing through that point. If there is a straight line artifact $y = m_0 x + c_0$ in the image, mapping each point (x, y) along the line will produce sets $\{(m, c)\}$ that intersect at the true parameters (m_0, c_0). These intersections can be found by binning the mapped sets $\{(m, c)\}$ over a grid in (m, c), identifying bins with the highest counts.

The Hough transform has been applied primarily to all-sky CW searches. As a two-dimensional transform, it requires selecting two dimensions from the at least four-dimensional all-sky parameter space. This has led to two implementations: SkyHough [158], which applies the Hough transform to the sky parameter space at fixed frequency and spindown; and FrequencyHough [159], which applies it to the frequency-spindown parameter space at fixed sky position. Aside from this choice, both implementations operate similarly.

A Hough transform pipeline begins with matched filter results from the first coherent stage for each segment, typically using power. Each matched filter result must surpass a threshold and be a local maximum relative to neighboring filters, selecting a collection of points known as a *peak map*. The Hough transform is then applied to the peak map, and its results are binned into a grid called a *partial Hough map*; each grid cell is one if it contains a value of the transform, or zero otherwise. Finally, the remaining dimensions of the search parameter space are considered (frequency and spindown for SkyHough, sky position for FrequencyHough). For each vector of parameters selected from these dimensions, the partial Hough maps consistent with a CW signal with these parameters are summed over segments to produce a *total Hough map*. The final output of the SkyHough (FrequencyHough) pipelines are total Hough maps over the sky (frequency-spindown) parameter spaces, for fixed values of frequency and spin-down (sky position). The detection statistic in each bin of the total Hough map is the count of segments (out of N) in which the corresponding bin of the partial Hough maps registered a one.

Further developments of each implementation have included: weighted summing of the partial Hough maps to account for the amplitude modulation of the CW signal [160]; BinarySkyHough, an extension of the SkyHough implementation to efficiently search over binary orbital parameters [161]; extensions of the FrequencyHough implementation to perform directed searches (i.e., at fixed sky position) efficiently using the BSD framework [162], and to search over binary orbital parameters [163]; and adaptations of

the Hough transform beyond the CW realm to search for long-duration but transient gravitational wave signals [164].

2.4.2.3 StackSlide algorithms

The StackSlide algorithm [165] operates as follows: Start by computing the power spectra from a set of SFTs, stacking each spectrum as a vertical column along a horizontal time axis to create a time-frequency plane of power versus time. A CW signal in this plane would appear as a wavy horizontal line due to various frequency modulations $f(t)$. By "sliding" each spectrum up and down for a given vector of sky and spin-down parameters, the wavy line can be straightened. Summing the frequency bins of the spectra in their adjusted positions accumulates CW signal power over time for a range of frequencies. The summed StackSlide power follows a χ^2 distribution with $2N$ degrees of freedom [165].

The StackSlide concept has been generalized to use the $\mathcal{F}$-statistic instead of power. The idea of "sliding" power spectra is concretely implemented using coarse and fine grids, as follows. First, in the coherent stage, the $\mathcal{F}$-statistic is computed over a coarse grid for each segment. This grid is constructed using the parameter-space metric (Eq. (58)) to decide grid spacings. In the semi-coherent stage, a fine grid is constructed to suit an incoherent sum of $\mathcal{F}$-statistic values. This is achieved by averaging the N coherent metrics from each segment. The fine grid typically contains γ more points per unit parameter space volume than the coarse grid; γ is the *refinement factor*.

Next, each coarse grid is interpolated onto the fine grid: for each fine grid point, and for each segment, the closest coarse grid point is identified (using the coherent parameter-space metric). This provides a one-to-N mapping from each fine grid point to its closest coarse grid points in each segment. For each fine grid point, the N values of $2\mathcal{F}$ from the closest coarse grid points are summed. The output detection statistic follows a χ^2 distribution with $4N$ degrees of freedom [165]. Implementing the StackSlide algorithm with the $\mathcal{F}$-statistic requires knowledge of both the coherent and semi-coherent metrics. A desirable property is for the metrics to be flat, i.e., $g_{ij}(\vec{p}) = g_{ij}$ is constant with respect to phase parameters, which aids in generating template banks that minimize computational cost [166]. However, such template banks may not be optimal for detection [167].

The metric is a quadratic approximation of signal power loss, valid for small mismatches $\mu \lesssim 0.4$ [127]. Due to computational constraints, the semi-coherent mismatch rarely meets this requirement, but empirical studies indicate that signal power loss progresses slowly even at large mismatches. Hence, StackSlide searches can still achieve good sensitivity, even with a large ($\gtrsim 1$) metric mismatch μ [168]. Building on large-scale correlations in phase parameter space and simplified CW phase models, the Global Correlation Transform (GCT) was derived in [169]. The GCT introduces new phase parameters by absorbing a Taylor expansion of Earth's orbital motion into new frequency and spin-down coordinates. The GCT metric underestimates signal power loss over realistic observing times but remains in use for all-sky CW searches. The supersky metric [125] expands on GCT ideas, addressing its limitations. It projects the sky parameter space from a 2-sphere to three dimensions, then back to a two-dimensional plane corresponding to the equatorial plane (short T) or ecliptic plane (long T). This retains higher-order terms

important for longer T. Combined with an optimal lattice-based template bank [170], the StackSlide search pipeline based on the supersky metric, known as Weave [171], has demonstrated improved sensitivity compared to GCT [172]. Weave has been used for narrow-band searches for known pulsars and directed searches for CCOs, despite its memory requirements for storing $2\mathcal{F}$ values for the fine-to-coarse grid mapping [172].

2.4.2.4 PowerFlux algorithms

PowerFlux [122] is an all-sky search pipeline for isolated neutron stars. The primary distinction is that PowerFlux weights the power from each SFT by the detector response functions, emphasizing times during the day when the detector is most sensitive for a given sky position. Additionally, PowerFlux inversely weights the power from each SFT by its noise, de-weighting times when detector sensitivity is degraded. The method is highly optimized [173] and often used for "quick-look" searches of early data from a run, capitalizing on the typical improvement in detector sensitivity after upgrades and commissioning between runs. PowerFlux has developed a procedure for computing h_0^{C} upper limits that are strictly conservative (i.e., worst-case) even in the presence of spectral artifacts [174]. The loosely coherent extension to the classic PowerFlux algorithm [175] generalizes the concept of summing SFT power over time. Instead, it considers a two-dimensional summation over all pairs of SFTs at times t_1, t_2, where a kernel function $K_{t_1,t_2}(\delta)$ with parameter δ decides which pairs to add and with what weight.

In this framework, summation of SFT power corresponds to a kernel that is one when $t_1 = t_2$ and zero otherwise. This kernel blurs the distinction between the coherent and semi-coherent stages in the hierarchical search paradigm, allowing a smoother transition between full phase coherence and phase-incoherent power summing over a timescale determined by δ. It bears similarities with the cross-correlation method (Sec. 2.4.2.5) and the approaches of [176]. The loosely coherent method was initially developed to follow up candidates from an initial PowerFlux search, enabling longer effective coherence times T_{coh} to be used, as seen in Fig. 2.5. A recent, fast implementation of the method called Falcon [177] enables loose coherence to be used in the initial search as well.

2.4.2.5 Cross-correlation algorithms

Cross-correlation is a well-established concept in signal processing, used to quantify the similarity between two independent time series as a function of their relative time offset. Its initial application to GW data was as a radiometer in software [178] for detecting unmodeled stochastic gravitational waves. Subsequently, a model-based cross-correlation algorithm was developed for searching for CW signals [179]. Similar to the loosely coherent PowerFlux kernel, each pair of independent Short Fourier Transforms (SFTs), labelled I and J—either from different times or different detectors—are cross-correlated using a filter Q_{IJ}. This filter weights each pair of SFTs according to how a CW signal would manifest in both, demodulating the signal to maximize the signal-to-noise ratio. The filter can be adjusted to select

which SFTs to cross-correlate: correlating only those close in time yields a power-like detection statistic, whereas correlating all SFTs recovers the $\mathcal{F}$-statistic. Thus, the cross-correlation method allows for flexible tuning of phase coherence, more so than traditional two-stage hierarchical searches.

The initial CrossCorr implementation targeted a CCO in the supernova remnant SN 1987A [180]. Further advancements include "resampling" for efficient frequency computations, similar to the $\mathcal{F}$-statistic. Lattice template placement has been employed to minimize computational cost, particularly by utilizing a coordinate transform in the $P - t_{\text{asc}}$ space, reducing the template bank to a single point in P [181]. Typically, CrossCorr adjusts its effective coherence length T_{max} based on parameters such as f, a_p, and t_{asc}.

2.4.2.6 TwoSpect algorithm

TwoSpect [182] is an algorithm specifically designed to detect CW signals from neutron stars with binary companions. The process begins similarly to the StackSlide-on-power method by stacking Short Fourier Transform (SFT) power spectra to create a time-frequency plane, where the vertical axis represents SFT frequency and the horizontal axis represents time. The Fourier transformations and power spectra are then computed over this time-frequency plane, resulting in a frequency-frequency plane where the vertical axis shows SFT frequency and the horizontal axis shows the second Fourier transform frequency, as illustrated in Fig. 1 of [182]. Due to the modulations of the CW signal from the orbits of the neutron star and Earth, signal power appears in the TwoSpect frequency-frequency plane at regularly spaced intervals. In the initial analysis stage, power in pixels and their harmonics is incoherently summed to identify promising candidates. The second stage then constructs templates that match the expected pattern of pixels for a CW signal with specified sky, frequency, and binary orbital parameters. TwoSpect has been used in directed search mode to target low-mass X-ray binaries (LMXBs) such as Sco X-1 and XTE J1751–305 [183].

2.4.2.7 Viterbi and SOAP algorithms

An HMM (Sec. 2.4.1.4) models the CW signal frequency as a randomly-changing path over a time-frequency plane, rather than as a pre-determined well-modeled function. The Viterbi algorithm is employed to determine the most likely path of the signal frequency through this plane.

The Viterbi algorithm is often described as "tracking" the signal forward in time, but it actually works by analyzing the data backwards in time. For CW searches [184], the Viterbi algorithm operates as follows: for each time step m and frequency bin n, it evaluates the frequency bins $n - \delta n_-, n - \delta n_+ 1, \ldots, n + \delta n_+$ from the previous time step $m - 1$. The values δn_- and δn_+ are chosen based on the expected behavior of the CW frequency, for instance, whether it may decrease ($\delta n_- > 0$) or increase ($\delta n_+ > 0$) over time. Based on the detection statistics computed at these δn frequency bins, the algorithm selects the bin with the highest detection statistic, indicating the most probable path from time step $m - 1$ to time step m. The maximum detection statistic at time step $m-1$ is then added to the detection statistic at the current

time step m and frequency bin n. In essence, at each time step m, the Viterbi algorithm identifies the N most likely past paths of the CW frequency that intersect the N frequency bins. By the final time step $M-1$, it has found the N most probable paths of the CW frequency through the time-frequency plane ending at the N frequency bins.

The Viterbi algorithm offers several advantages, including good coverage of a wide range of possible signal frequency variations, low computational cost, and adaptability to various search targets. The first implementation for CW searches [143] focused on LMXBs like Sco X-1. Unlike templated CW algorithms, the Viterbi algorithm effectively handles the expected spin wandering of the signal frequency due to time-varying accretion torque. This approach has also been applied to CCOs in young supernova remnants [185] and long-duration transient gravitational waves from neutron stars formed in binary neutron star mergers.

A variant, SOAP [184], enhances the basic Viterbi algorithm by incorporating "memory", allowing the algorithm to consider multiple previous time steps rather than just one, improving the search for periodic-like signals. A post-processing stage using convolutional neural networks [186] enhances robustness against spectral artefacts. SOAP is designed as a versatile, "quick-look" search method for CW signals and for identifying instrumental line artefacts.

2.4.2.8 Machine learning

The application of machine-learning (ML) techniques in the search for CW signals typically follows one of two approaches: either classifying and summarizing the results of a search's main stage, or substituting the search step itself by acting as a detection statistic. The first approach involves training a classifier directly on Fourier-transformed raw time-series data to distinguish the presence of a signal within background noise. The specific data format depends on the type of signals being targeted: for CW signals, which last for long periods over narrow frequency bands, a convolutional neural network (CNN) can be trained on the real and imaginary parts of short-time Fourier transforms [130]. Variations of this method have been used to search for post-merger signals, which, in contrast to CW signals, have shorter durations and cover wider frequency bands [187]. To account for the non-trivial time dependency of the signal, Ref. [131] utilized shorter time durations for Fourier transforms and used the resulting spectrogram data as input for a CNN. The second approach is less radical and involves applying ML algorithms to the output of a search pipeline to construct a new detection statistic [133, 139]. This approach can be advantageous, as the output of search pipelines typically enhances signal features across the parameter space. This thesis focuses on the development of an algorithm that utilizes advanced ML techniques to detect signals within Fourier-transformed data streams. Specifically, the algorithm processes SFT data to produce a detection statistic indicating the presence or absence of a signal. The following chapter details the implementation of this approach.

Chapter 3

Deep Learning

This chapter introduces the fundamental concepts of deep learning and its applications to CW detection. It specifically highlights the central role of Convolutional Neural Networks (CNNs) and the technique of Transfer Learning in this context. Then, chapter 4 addresses the adaptation of such methods to GW physics and the development of a Binary classifier capable of detecting CW signals embedded in Gaussian noise and Real detector data.

3.1 Feed-Forward Deep Neural Networks (DNNs)

Over the past decade, neural networks have become one of the most powerful and widely-used techniques for supervised learning. Deep Neural Networks (DNNs) have a long history [188], but gained renewed prominence in the mid-2000s with the rebranding as "Deep Learning" [189]. The adoption of DNNs by the broader machine learning community and industry was notably accelerated in 2012 with the advent of GPU-based DNN model (AlexNet)[190] to significantly reduce the error rate in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) from 26% to 16% [190]. This success was followed three years later by Microsoft achieving an error rate of 3.57% using an ultra-deep residual neural network (ResNet) with 152 layers [191]. Since then, DNNs have become essential tools for many image and speech recognition-based machine learning tasks. The large-scale industrial deployment of DNNs has led to the creation of many high-level libraries and packages (Keras [192], Pytorch [193], TensorFlow [194], etc.), making it easier to quickly develop and deploy DNNs.

Conceptually, neural networks can be divided into four categories: (i) general-purpose neural networks for supervised learning, (ii) neural networks designed specifically for image processing, with Convolutional Neural Networks (CNNs) being the most prominent example, (iii) neural networks for sequential data such as Recurrent Neural Networks (RNNs), and (iv) neural networks for unsupervised learning such as Deep Boltzmann Machines. The first two categories (supervised learning and image processing), are the most widely used and the two types used for the research subject in this thesis.

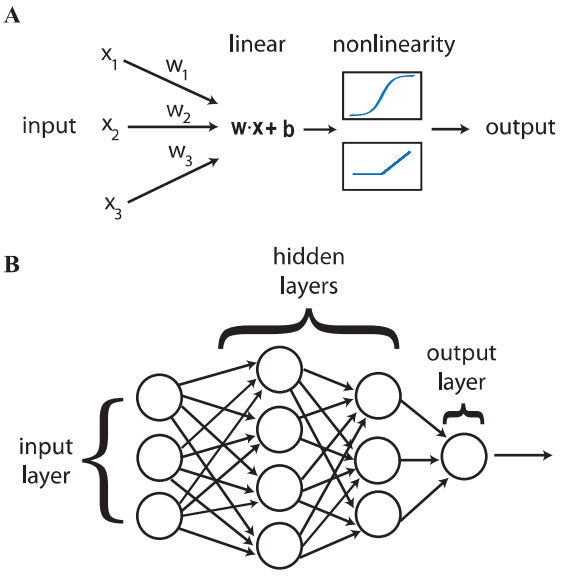


FIGURE 3.1: Basic architecture of neural networks. (A) The basic components of a neural network are stylized neurons consisting of a linear transformation that weights the importance of various inputs, followed by a non-linear activation function. (B) Neurons are arranged into layers with the output of one layer serving as the input to the next layer

3.1.1 Neurons

Neural networks are inspired by the structure of the real brain neurons and are used for supervised learning tasks. The basic unit of a neural net is a "neuron" that takes a vector of d input features $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)$ and produces a scalar output $a_i(\boldsymbol{x})$. A neural network consists of many such neurons stacked into layers, with the output of one layer serving as the input for the next, as shown in Fig. 3.1. The first layer in the neural net is called the input layer, the middle layers are often called "hidden layers", and the final layer is called the output layer. The exact function a_i varies depending on the type of non-linearity used in the neural network. However, a_i can be decomposed into a linear operation that weights the relative importance of the various inputs, and a non-linear transformation $\sigma_i(z)$ which is usually the same for all neurons in the network. The linear transformation takes the form of a dot product with a set of neuron-specific weights $\boldsymbol{w}^{(i)} = (w_1^{(i)}, w_2^{(i)}, \ldots, w_d^{(i)})$ followed by re-centering with a neuron-specific bias $b^{(i)}$:

$$z^{(i)} = \boldsymbol{w}^{(i)} \cdot \boldsymbol{x} + b^{(i)} = \boldsymbol{x}^T \cdot \boldsymbol{w}^{(i)}, \tag{3.1}$$

where $\boldsymbol{x} = (1, x)$ and $\boldsymbol{w}^{(i)} = (b^{(i)}, \boldsymbol{w}^{(i)})$. In terms of $z^{(i)}$ and the non-linear function $\sigma_i(z)$, the full input-output function is:

$$a_i(\boldsymbol{x}) = \sigma_i(z^{(i)}). \tag{3.2}$$

Historically, common choices of nonlinearities included step-functions (perceptrons), sigmoids (i.e., Fermi functions), and the hyperbolic tangent. Recently, it has become more common to use rectified linear units (ReLUs) [195], leaky rectified linear units (leaky ReLUs), and exponential linear units (ELUs), as depicted in Fig. 3.2. Different choices of nonlinearities lead to different computational and training properties. The underlying reason for this is that neural nets are trained using gradient descent based

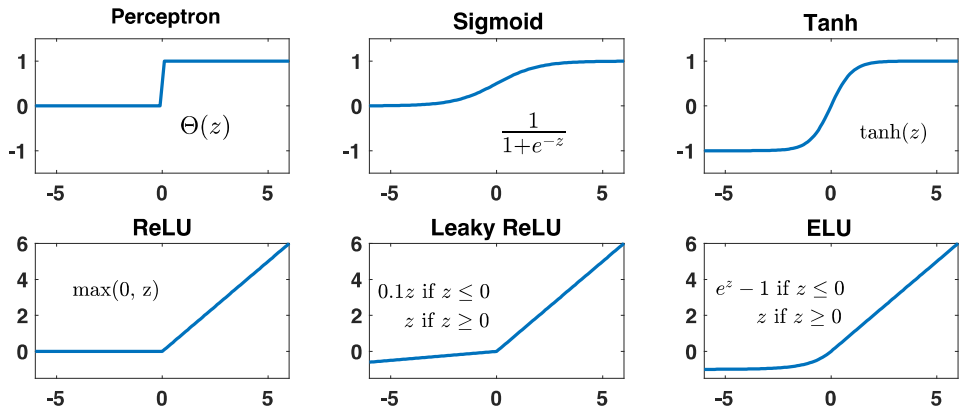


FIGURE 3.2: Possible non-linear activation functions for neurons. In modern DNNs, it has become common to use non-linear functions that do not saturate for large inputs (bottom row) rather than saturating functions (top row).

methods, that require us to take derivatives of the neural input-output function with respect to the weights $\boldsymbol{w}^{(i)}$ and the bias $b^{(i)}$.

Until recently the most popular choice of non-linearity was the tanh function or a sigmoid/Fermi function. However, when the input weights become large, as they often do in training, the activation function saturates and the derivative of the output with respect to the weights tends to zero since $\partial\sigma/\partial z \to 0$ for $z \gg 1$. Such "vanishing gradients" are a feature of any saturating activation function (top row of Fig. 3.2), making it harder to train deep networks. In contrast, for a non-saturating activation function such as ReLUs or ELUs, the gradients stay finite even for large inputs.

3.1.2 Network architecture

Neural networks are a collection of layer neurons stacked in a hierarchical fashion, the general structure of which is known as the network architecture in Fig. 3.1. In the simplest feed-forward networks, each neuron in the input layer of the neurons takes the inputs x and produces an output $a_i(x)$ that depends on its current weights (Eq. (3.2)). The outputs of the input layer are then treated as the inputs to the next hidden layer. This is repeated several times until one reaches the top or output layer. The output layer is a simple classifier like a logistic regression or soft-max function in the case of categorical data (i.e. discrete labels) or a linear regression layer in the case of continuous outputs. Thus, the whole neural network can be thought of as a complicated nonlinear transformation of the inputs x into an output $\hat{y}$ that depends on the weights and biases of all the neurons in the input, hidden, and output layers.

The use of hidden layers greatly expands the representational power of a neural net when compared with a simple soft-max or linear regression network. Formally, this is a result of the universal approximation theorem [196] which states that a neural network with a single hidden layer can approximate any continuous, multi-input/multi-output function with arbitrary accuracy. In other words, hidden neurons allow neural networks to generate step functions with arbitrary offsets and heights. These can then be added together to approximate arbitrary functions. The more complicated a function, the more hidden

units (and free parameters) are needed to approximate it. Hence, the applicability of the approximation theorem to practical situations should not be overemphasized.

Modern neural networks generally contain multiple hidden layers (hence the "deep" in deep learning). There are many ideas of why such deep architectures are favorable for learning. Increasing the number of layers increases the number of parameters and hence the representational power of neural networks. Indeed, recent numerical experiments suggest that as long as the number of parameters is larger than the number of data points, certain classes of neural networks can fit arbitrarily labeled random noise samples [197].

Choosing the exact network architecture for a neural network remains an "art" that requires extensive numerical experimentation and intuition and is oftentimes problem-specific. Both the number of hidden layers and the number of neurons in each layer can affect the performance of a neural network. There is no single recipe for the right architecture for a neural net that works best. However, a consensus that seems to be emerging is that the number of parameters in the neural net should be large enough to prevent under-fitting (see theoretical discussion in [198]). Empirically, the best architecture for a problem depends on the task, the amount and type of data that is available, and the computational resources at one's disposal. Certain architectures are easier to train, while others might be better at capturing complicated dependencies in the data and learning relevant input features. Finally, there have been numerous works that move beyond the simple deep, feed-forward neural network architectures discussed here. For example, modern neural networks for image segmentation often incorporate "skip connections" that skip layers of the neural network [191]. This allows information to directly propagate to a hidden or output layer, bypassing intermediate layers and often improving performance.

3.1.3 Training process

The fundamental procedure for training neural networks mirrors that of simpler supervised learning algorithms like logistic and linear regression: construct a cost/loss function and use gradient descent to minimize it, thereby finding the optimal weights and biases. Neural networks differ from these simpler supervised procedures in that they generally contain multiple hidden layers, making the gradient computation more complex. The "backpropagation" algorithm [199] for computing gradients, addresses this complexity.

As with all supervised learning procedures, the first step in training a neural network is to specify a loss function. Given a data point $(\boldsymbol{x}_i, y_i)$ where $\boldsymbol{x}_i \in \mathbb{R}^{d+1}$, the neural network makes a prediction $\hat{y}_i(\boldsymbol{w})$, where $\boldsymbol{w}$ represents the network parameters. The output layer typically serves as either a continuous predictor or a classifier. For continuous predictions, common loss functions include the mean squared error (MSE)

$$E(\boldsymbol{w}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i(\boldsymbol{w}))^2, \tag{3.3}$$

where n is the number of data points, and the mean-absolute error (i.e., L_1 norm)

$$E(\boldsymbol{w}) = \frac{1}{n}\sum_i |y_i - \hat{y}_i(\boldsymbol{w})|. \tag{3.4}$$

The full cost function often includes additional terms for regularization (e.g., L_1 or L_2 regularizers (**ref to regularizers**). For categorical data, the cross-entropy loss function is commonly used, especially when the output layer acts as a logistic or softmax classifier. The cross-entropy between true labels $y_i \in \{0, 1\}$ and predictions is given by

$$E(\boldsymbol{w}) = -\sum_{i=1}^{n} y_i \log \hat{y}_i(\boldsymbol{w}) + (1 - y_i) \log[1 - \hat{y}_i(\boldsymbol{w})]. \tag{3.5}$$

For categorical data with M possible classes, the categorical cross-entropy is defined as

$$E(\boldsymbol{w}) = -\sum_{i=1}^{n}\sum_{m=0}^{M-1} y_{im} \log \hat{y}_{im}(\boldsymbol{w}) + (1 - y_{im}) \log[1 - \hat{y}_{im}(\boldsymbol{w})], \tag{3.6}$$

where $y_{im} = 1$ if $y_i = m$ and 0 otherwise. After defining the architecture and cost function, the next step is to train the model using gradient-based methods to optimize the cost function. The purpose of gradient descent is to update the parameters in the direction of the negative gradient of the cost function $\nabla_w E(\boldsymbol{w})$. Modern neural network packages, such as Keras [192], allow the specification of these optimizers. Depending on the architecture, data, and computational resources, different optimizers may perform better, though standard SGD is often a good starting point. Calculating gradients for a neural network requires the specialized backpropagation algorithm, which is central to any neural network training procedure.

3.1.4 The Back-propagation Algorithm

The back-propagation algorithm [200] leverages the layered structure of neural networks to efficiently compute gradients. At its core, back propagation is the application of the ordinary chain rule for partial differentiation. Consider a network with L layers, indexed by $l = 1, \ldots, L$. Denote by w_{jk}^l the weight for the connection from the k-th neuron in layer $l-1$ to the j-th neuron in layer l. The bias of this neuron is represented by b_j^l. In a feed-forward neural network, the activation a_j^l of the j-th neuron in the l-th layer is related to the activations of the neurons in the $l-1$-th layer by

$$a_j^l = \sigma\left(\sum_k w_{jk}^l a_k^{l-1} + b_j^l\right) = \sigma(z_j^l), \tag{3.7}$$

where the linear weighted sum z_j^l is defined as

$$z_j^l = \sum_k w_{jk}^l a_k^{l-1} + b_j^l. \tag{3.8}$$

The cost function E directly depends on the activities of the output layer a_j^L and indirectly on the activities of neurons in the lower layers via iteration of the Eq. (3.8). Define the error Δ_j^L of the j-th neuron in the L-th layer as the change in the cost function with respect to the weighted input z_j^L

$$\Delta_j^L = \frac{\partial E}{\partial z_j^L}. \tag{3.9}$$

Analogously, the error of neuron j in layer l, Δ_j^l, is defined as the change in the cost function with respect to the weighted input z_j^l

$$\Delta_j^l = \frac{\partial E}{\partial z_j^l} = \frac{\partial E}{\partial a_j^l}\sigma'(z_j^l), \tag{3.10}$$

where $\sigma'(x)$ denotes the derivative of the non-linearity $\sigma(\cdot)$ with respect to its input evaluated at x. The error function Δ_j^l can also be interpreted as the partial derivative of the cost function with respect to the bias b_j^l, since

$$\Delta_j^l = \frac{\partial E}{\partial z_j^l} = \frac{\partial E}{\partial b_j^l}\frac{\partial b_j^l}{\partial z_j^l} = \frac{\partial E}{\partial b_j^l}, \tag{3.11}$$

where the fact that $\partial b_j^l / \partial z_j^l = 1$ has been used. Now, consider that the error depends on neurons in layer l only through the activation of neurons in the subsequent layer $l+1$. The chain rule allows writing

$$\begin{aligned}\Delta_j^l = \frac{\partial E}{\partial z_j^l} &= \sum_k \frac{\partial E}{\partial z_k^{l+1}}\frac{\partial z_k^{l+1}}{\partial z_j^l} \\ &= \sum_k \Delta_k^{l+1}\frac{\partial z_k^{l+1}}{\partial z_j^l} \\ &= \left(\sum_k \Delta_k^{l+1} w_{kj}^{l+1}\right)\sigma'(z_j^l).\end{aligned} \tag{3.12}$$

Next, by differentiating the cost function with respect to the weight w_{jk}^l as

$$\frac{\partial E}{\partial w_{jk}^l} = \frac{\partial E}{\partial z_j^l}\frac{\partial z_j^l}{\partial w_{jk}^l} = \Delta_j^l a_k^{l-1}. \tag{3.13}$$

The basic equations for backpropagation Eqs. (3.10), (3.11), (3.12), and (3.13) define the four backpropagation equations relating the gradients of the activations of various neurons a_j^l, the weighted inputs $z_j^l = \sum_k w_{jk}^l a_k^{l-1} + b_j^l$, and the errors Δ_j^l. These equations can be combined into a simple, computationally efficient algorithm to calculate the gradient with respect to all parameters [199]. In summary, the algorithm works as follows:

1. **Activation at input layer**: Calculate the activations a_j^1 of all the neurons in the input layer.
2. **Feedforward**: Starting with the first layer, exploit the feedforward architecture through Eq. (3.7) to compute z^l and a^l for each subsequent layer.
3. **Error at top layer**: Calculate the error of the top layer using Eq. (3.10). This requires knowing the expression for the derivative of both the cost function $E(\boldsymbol{w}) = E(a^L)$ and the activation function $\sigma(z)$.

4. **Back-propagate the error**: Use Eq. (3.12) to propagate the error backwards and calculate Δ_j^l for all layers.

5. **Calculate gradient**: Use Eqs. (3.11) and (3.13) to calculate $\partial E/\partial b_j^l$ and $\partial E/\partial w_{jk}^l$.

The name "back-propagation" is derived from the algorithm's process: a forward pass from the bottom layer to the top layer calculates the weighted inputs and activations of all neurons. The error is then back-propagated from the top layer to the input layer, using these errors to calculate the desired gradients. Computational efficiency is crucial since it is necessary to calculate the gradient with respect to all parameters of the neural network at each step of gradient descent. These principles underlie most modern automatic differentiation packages such as Autograd like PyTorch [193].

3.1.4.1 Potential Issues with Back-propagation

Equipped with back-propagation and gradient descent, the training of neural networks may initially appear straightforward. However, back-propagation combined with gradient descent on large networks is extremely computationally expensive. Advances in computational hardware, particularly the widespread use of GPUs, have mitigated this issue, making it less complex than it was a decade ago. On a more technical note, a prevalent issue in deep networks is the problem of *vanishing or exploding gradients*. This issue is especially pronounced in neural networks designed to capture long-range dependencies, such as Recurrent Neural Networks (RNNs) for sequential data. To illustrate this problem, consider a simple network with a single neuron in each layer and assume all weights are equal, denoted by w. The behavior of the back-propagation equations for such a network can be inferred by repeatedly applying Eq. (3.12):

$$\Delta_j^l = \Delta_j^L \prod_{j=0}^{L-1} w'\sigma'(z_j) = \Delta_j^L (w')^{L-1}\sigma'(z_j), \tag{3.14}$$

where Δ_j^L is the error in the L-th topmost layer, and $(w')^L$ represents the weight raised to the power L. Assuming the magnitude of $\sigma'(z_j)$ is fairly constant and approximating $\sigma'(z_j) \approx \sigma'$, it is evident that for large L, the error Δ_j^l exhibits significantly different behavior depending on the value of $w'\sigma'$. If $w'\sigma' > 1$, the errors and gradients escalate. Conversely, if $w'\sigma' < 1$, the errors and gradients diminish. Only when the weights satisfy $w'\sigma' \approx 1$ and the neurons are not saturated will the gradient remain stable for deep networks.

This fundamental behavior persists even in more complex networks. Rather than considering a single weight, it is essential to examine the eigenvalues (or singular values) of the weight matrices w_{jk}^l. To ensure gradients remain finite in deep networks, these eigenvalues must stay near unity after many gradient descent steps. In modern feed-forward and ReLU neural networks, this is achieved by initializing the weights such that the gradient does not vanish or explode. Employing non-linearities that do not saturate, such as ReLUs (noting that for saturating functions, $\sigma' \to 0$, causing the gradient to vanish), is another effective strategy. Proper initialization and regularization methods, such as gradient clipping (limiting

the magnitude of gradients) and batch normalization, also help mitigate the vanishing and exploding gradient problem.

3.1.5 Regularizing Neural Networks and Other Practical Considerations

DNNs, like all supervised learning algorithms, must navigate the bias-variance trade-off [201]. Regularization techniques play a significant role in ensuring that DNNs generalize well to new data. The past ten years have seen a proliferation of specialized regularization techniques for DNNs beyond the simple L_1 and L_2 penalties [202]. Techniques such as Dropout and Batch Normalization have become prevalent. In addition to these, large DNNs benefit from implicit regularization that naturally occurs in Stochastic Gradient Descent (SGD) [203]. The inherent stochasticity and local nature of SGD often prevent over-fitting on spurious correlations in training data, particularly when combined with methods like Early Stopping. This section provides a brief overview of these regularization techniques.

3.1.5.1 Stochastic Gradient Descent (SGD) with Mini-batches

One of the most widely-applied variants of the gradient descent algorithm is stochastic gradient descent (SGD) [204, 205]. Unlike ordinary gradient descent, the SGD algorithm is stochastic, incorporating stochasticity by approximating the gradient on a subset of the data, known as a *minibatch*. The size of the minibatches is typically much smaller than the total number of data points n, with common minibatch sizes ranging from ten to a few hundred data points. Given n points in total, and a minibatch size M, there will be n/M minibatches. Denote these minibatches by B_k where $k = 1, \ldots, n/M$. Thus, in SGD, at each gradient descent step, the gradient is approximated using a single minibatch B_k:

$$\nabla_\theta E(\boldsymbol{\theta}) = \sum_{i=1}^{n} \nabla_\theta e_i(\boldsymbol{x}_i, \boldsymbol{\theta}) \rightarrow \sum_{i \in B_k} \nabla_\theta e_i(\boldsymbol{x}_i, \boldsymbol{\theta}). \tag{3.15}$$

All $k = 1, \ldots, n/M$ minibatches are cycled through one at a time, using the minibatch approximation to update the parameters θ at each step k. A full iteration over all n data points, utilizing all n/M minibatches, is called an epoch. For notational convenience, the minibatch approximation to the gradient is denoted by

$$\nabla_\theta E^{MB}(\boldsymbol{\theta}) = \sum_{i \in B_k} \nabla_\theta e_i(\boldsymbol{x}_i, \boldsymbol{\theta}). \tag{3.16}$$

Using this notation, the SGD algorithm can be expressed as:

$$\boldsymbol{v}_t = \eta_t \nabla_\theta E^{MB}(\boldsymbol{\theta}), \quad \boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \boldsymbol{v}_t. \tag{3.17}$$

In SGD, the actual gradient over the full data at each gradient descent step is replaced by an approximation using a minibatch. This approach provides two significant benefits: introducing stochasticity reduces the likelihood of the fitting algorithm becoming trapped in isolated local minima, and it substantially speeds up the calculation by eliminating the need to use all n data points to approximate the gradient.

Empirical and theoretical research indicates that SGD offers additional advantages, such as acting as a natural regularizer that mitigates over-fitting in deep, isolated minima [206, 207].

3.1.5.2 Adding Momentum

In practice, Stochastic Gradient Descent (SGD) often includes a "momentum" or inertia term that retains the memory of the direction in parameter space. This is typically implemented as follows:

$$\begin{aligned} \boldsymbol{v}_t &= \gamma \boldsymbol{v}_{t-1} + \eta_t \nabla_\theta E(\boldsymbol{\theta}_t) \\ \boldsymbol{\theta}_{t+1} &= \boldsymbol{\theta}_t - \boldsymbol{v}_t, \end{aligned} \tag{3.18}$$

where the momentum parameter γ (with $0 \leq \gamma \leq 1$) is introduced. The notation indicating that the gradient is taken over a different mini-batch at each step is omitted for brevity. This method, known as gradient descent with momentum (GDM), reveals that v_t represents a running average of recently encountered gradients, and $(1-\gamma)^{-1}$ defines the characteristic time scale for the memory used in the averaging procedure. When $\gamma = 0$, this reduces to ordinary SGD as described in Eq. (3.17). An equivalent way of expressing the updates is

$$\Delta \boldsymbol{\theta}_{t+1} = \gamma \Delta \boldsymbol{\theta}_t - \eta_t \nabla_\theta E(\boldsymbol{\theta}_t), \tag{3.19}$$

with $\Delta\theta_t = \theta_t - \theta_{t-1}$. Momentum-based methods were first introduced in early, largely overlooked Soviet research [208]. To provide intuition, consider a physical analogy with a particle of mass m moving in a viscous medium with damping coefficient μ and potential $E(\boldsymbol{w})$ [209]. If $\boldsymbol{w}$ denotes the particle's position, its motion is governed by

$$m\frac{d^2\boldsymbol{w}}{dt^2} + \mu\frac{d\boldsymbol{w}}{dt} = -\nabla_w E(\boldsymbol{w}). \tag{3.20}$$

Discretization of this equation yields

$$m\frac{\boldsymbol{w}_{t+\Delta t} - 2\boldsymbol{w}_t + \boldsymbol{w}_{t-\Delta t}}{(\Delta t)^2} + \mu\frac{\boldsymbol{w}_{t+\Delta t} - \boldsymbol{w}_t}{\Delta t} = -\nabla_w E(\boldsymbol{w}). \tag{3.21}$$

Rearranging gives

$$\Delta \boldsymbol{w}_{t+\Delta t} = -\frac{(\Delta t)^2}{m + \mu \Delta t}\nabla_w E(\boldsymbol{w}) + \frac{m}{m + \mu \Delta t}\Delta \boldsymbol{w}_t. \tag{3.22}$$

This equation is identical to Eq. (3.19) if the particle's position $\boldsymbol{w}$ is identified with the parameters θ. This identification allows the momentum parameter and learning rate to be expressed in terms of the particle's mass and viscous damping as

$$\gamma = \frac{m}{m + \mu \Delta t}, \quad \eta = \frac{(\Delta t)^2}{m + \mu \Delta t}. \tag{3.23}$$

Thus, the momentum parameter is proportional to the particle's mass, providing inertia. In the high-viscosity/small learning rate limit, the memory time scales as $(1-\gamma)^{-1} \approx m/(\mu \Delta t)$.

Momentum in SGD helps the gradient descent algorithm accelerate in directions with persistent but small gradients, while suppressing oscillations in high-curvature directions. This is crucial when the landscape is shallow and flat in some directions and narrow and steep in others. First-order methods, with appropriate initial conditions, can perform comparably to more expensive second-order methods, especially in complex deep learning models [210]. Empirical studies indicate that incorporating momentum is particularly beneficial during the initial "transient phase" of training. Enhancements like Nesterov Accelerated Gradient (NAG) [208, 210] further improve momentum's effectiveness by calculating the gradient at the expected value of the parameters given the current momentum. The NAG update rule is

$$\begin{aligned} \boldsymbol{v}_t &= \gamma \boldsymbol{v}_{t-1} + \eta_t \nabla_\theta E(\boldsymbol{\theta}_t + \gamma \boldsymbol{v}_{t-1}) \\ \boldsymbol{\theta}_{t+1} &= \boldsymbol{\theta}_t - \boldsymbol{v}_t. \end{aligned} \tag{3.24}$$

NAG allows for a larger learning rate compared to GDM for the same choice of γ. Other alternative optimization schemes include:

- **Adagrad**. In the stochastic gradient descent and momentum methods, all parameters are updated with the same learning rate. However, using a different learning rate for each parameter can accelerate convergence. Based on this concept, Adagrad [211] introduces a technique where the learning rate is adjusted using the magnitude of the gradients of each parameter. The parameters are updated as follows:

$$v_t = \nabla E(\theta_t), \tag{3.25}$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{G^{(i)} + \epsilon}} \odot v_t^{(i)}, \tag{3.26}$$

  where ϵ is a small value to avoid division by zero, and $\odot$ is the Hadamard product. The j-th component of $G^{(i)}$ is defined as:

$$G_t^{(i)} = \sum_{k \leq i} (v_t^{(k)})^2. \tag{3.27}$$

- **RMSProp**. RMSProp [212] is another optimizer that uses an adaptive learning rate. The update equations are as follows

$$g_t = \nabla E(\theta_t), \tag{3.28}$$

$$v_t = \beta v_{t-1} + (1 - \beta) g_t \odot g_t, \tag{3.29}$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{v_t + \epsilon}} \odot g_t. \tag{3.30}$$

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

- **Adam**. Is one of the most popular algorithms due to its stability. It combines ideas from both the momentum and RMSProp methods [213]. The updating equations are

$$g_t = \nabla E(\theta_t, \tag{3.31}$$

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t, \tag{3.32}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t \odot g_t, \tag{3.33}$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^i}, \tag{3.34}$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^i}, \tag{3.35}$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \odot \hat{m}_t. \tag{3.36}$$

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

3.1.5.3 Dropout

Another significant regularization method widely adopted in neural network literature is Dropout [214]. The primary purpose of Dropout is to prevent over-fitting by reducing spurious correlations between neurons within the network. Dropout circumvents these issues by randomly dropping out neurons (along with their connections) from the neural network during each training step. Typically, in each mini-batch of the gradient descent step, a neuron is dropped from the network with a probability p. The gradient descent step is then performed only on the weights of the "thinned" network of individual predictors. During training, on average, weights are only present for a fraction p of the time. Predictions are made by re-weighting the weights by p: $\boldsymbol{w}_{\text{test}} = p\boldsymbol{w}_{\text{train}}$. The learned weights can be viewed as an "average" weight over all possible thinned neural networks.

3.1.5.4 Batch Normalization

The fundamental idea behind Batch Normalization is the long-recognized observation that training in neural networks is most effective when inputs are centered around zero with respect to their mean. This prevents neurons from saturating and gradients from vanishing in deep networks [215]. Without such centering, parameter changes in lower layers can propagate to higher layers, causing saturation effects and vanishing gradients.

Batch Normalization introduces additional "BatchNorm" layers that standardize inputs by the mean and variance of the mini-batch. Consider a layer l with d neurons whose inputs are $(z_1^l, \ldots, z_d^l)$. Each dimension is standardized so that:

$$z_k^l \to \hat{z}_k^l = \frac{z_k^l - \mathbb{E}[z_k^l]}{\sqrt{\text{Var}[z_k^l]}}, \tag{3.37}$$

where the mean and variance are computed over all samples in the mini-batch. However, this procedure might alter the representational power of the neural network, particularly for non-linearities like tanh, which could force the network to operate in the linear regime around $z = 0$. Since non-linearities are vital for the representational capacity of DNNs, this could significantly impact their performance. To mitigate this, Batch Normalization introduces two new parameters γ_k^l and β_k^l for each neuron, allowing the model to adjust the mean and variance:

$$\hat{z}_k^l \to \tilde{z}_k^l = \gamma_k^l \hat{z}_k^l + \beta_k^l. \tag{3.38}$$

These additional parameters are learned just like weights and biases using back-propagation. Initially, during training, the inputs are standardized, and back-propagation adjusts γ and β. In practice, Batch Normalization significantly improves learning speed by preventing gradient vanishing.

3.2 Convolutional Neural Networks (CNNs)

An important realization in Physics is the exploitation of symmetries and invariances when analyzing complex physical systems. Properties such as locality and translational invariance are often embedded directly into physical laws. Statistical physics models frequently incorporate all known information about the physical system under analysis. For instance, in many cases, it is sufficient to consider only local couplings in Hamiltonians or to work directly in momentum space if the system is translationally invariant. Tailoring analysis to exploit additional structure is a fundamental aspect of modern physical theories, from general relativity to gauge theories to critical phenomena.

Similarly, many datasets and supervised learning tasks in machine learning exhibit additional symmetries and structure. Consider a supervised learning task aimed at labeling images from a dataset as either containing cats or not. The statistical procedure must first learn features associated with cats. Given that a cat is a physical object, it is known that these features are likely to be local (groups of neighboring pixels in the two-dimensional image corresponding to tails, eyes, etc.). The location of these features within the image is less relevant, though their relative positions may matter, indicating translational invariance inherent to the supervised learning task.

The architecture previously discussed does not include nor consider this symmetry structure. For example, in the case of the digit 'four' from the MNIST dataset [216], the 28×28 image is treated as a one-dimensional vector of size $28^2 = 796$, discarding much of the spatial information. Recognizing this limitation, the neural network community developed convolutional neural networks (CNNs) that leverage this additional structure (locality and translational invariance) [217]. Interestingly, from a physics perspective, recent findings indicate that these CNN architectures are closely related to models such as tensor networks [218] and, specifically, MERA-like architectures used in quantum condensed matter systems [219].

3.2.1 CNNs Architecture

In contrast to regular neural networks, the layers of CNNs have neurons arranged in a few dimensions: channels, width, height, and number of filters in the simplest 2D case. A CNN consists, of a sequence of layers, where every layer transforms the activations or outputs of the previous layer through another differentiable function. There are several such layers employed in CNNs, however, the most common building blocks which you will encounter in most CNN architectures are: the convolution layer, pooling layer, and fully connected layers. In essence, these layers are like feature extractors, dimensionality reduction and classification layers, respectively. These layers of a CNN are stacked to form a full convolutional layer.

Essentially, a convolution layer uses a convolutional kernel as a filter for the input. Usually, there are many of such filters. During a forward pass, a filter slides across the input volume and computes the activation map of the filter at that point by computing the point-wise product of each value and adding these to obtain the activation at the point. Such a sliding filter is naturally implemented by a convolution and, as this is a linear operator, it can be written as a dot-product for efficient implementation.

3.2.1.1 Convolutions

Mathematically, the convolution $(x * w)(a)$ of functions x and w is defined in all dimensions as

$$(x * w)(a) = \int x(t)w(a-t)\,dt, \tag{3.39}$$

where a is in $\mathbb{R}^n$ for any $n \geq 1$, and the integral is replaced by its higher-dimensional variant. To understand the idea behind convolutions, consider the Gaussian function $w(a) = \exp(-a^2)$ as an example. If a photo is taken with a camera and the camera is slightly shaken, the blurry picture would be the real picture x convolved with a Gaussian function w. In the terminology of convolutional neural networks, x is called the input, w is called the filter or kernel, and the output is often referred to as activation or feature map. Note that in Eq. (3.39), the input and kernel are modeled as continuous functions. Due to the discrete nature of image sensors, in practice a discretized version is used. Then the discrete convolution can be defined as

$$(x * w)(a) = \sum_t x(t)w(t-a), \tag{3.40}$$

where a runs over all values in space and can be in any dimension. In deep learning, usually x is a multidimensional array of data and the kernel w involves learnable parameters and typically has finite support, meaning there are only finitely many values a for which $w(a)$ is nonzero. This allows Eq. (3.40) to be implemented as a finite summation. The definition of (3.40) is independent of dimension, but in general, 2- or 3-dimensional convolutions are mainly used:

$$(I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n), \tag{3.41}$$

or

$$(I * K)(i, j, k) = \sum_m \sum_n \sum_\ell I(m, n, \ell) K(i - m, j - n, k - \ell). \tag{3.42}$$

The convolutions (3.39) and (3.40) are commutative, meaning $I * K = K * I$, so (3.41) can be written as

$$(I * K)(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n). \tag{20.14}$$

As K has finite support, this a priori infinite sum becomes finite. Some neural network libraries also implement an operation called cross-correlation, but from a deep learning perspective, these operations are equivalent, as one weight set can be directly translated into the other.

As a convolution is a linear transformation, it can be written in the form of $w \cdot x + b$ and thus as a fully connected layer. However, the kernels are often much smaller than the input, and only a small number of inputs will interact with the output (the so-called receptive field), resulting in a sparse weight tensor w. Additionally, the weight tensor will contain many similar elements because the kernel is applied to every location in the input. This effect, referred to as weight sharing, along with the sparsity, implies that fewer parameters need to be stored, improving both memory requirements and statistical efficiency while placing a prior on the weights. It is implicitly assumed that a filter, such as an edge filter, is relevant to every part of the image and that most interactions between pixels are local.

Significant research has been conducted into adapting convolutions and imposing new strong priors, such as the group convolution [220], which enforces a certain symmetry group to hold for the image. In addition to sparsity and weight sharing, convolutions impose another prior on the kernel weights in the form of translation equivariance. Translation (or shifting) of the convolution implies that applying a translation to the image and then applying convolutions yields the same result as first applying the convolution and then translating the feature map. More specifically, an operator T is said to be equivariant with respect to f if for each x it holds that $T(f(x)) = f(T(x))$. Translation equivariance is a sensible assumption for images, as the features to detect an object in the image should only depend on the object itself and not on its precise location.

3.2.1.2 Pooling Layers

The goal of a pooling layer is to produce a summary statistic of its input and to reduce the spatial dimensions of the feature map while retaining essential information. Max pooling layers report the maximal values in each rectangular neighborhood of each point (i, j) (or (i, j, k) for 3D data) of each input feature, whereas average pooling layers report the average values. The most common form of max pooling uses stride 2 with kernel size 2, which partitions the feature map into a regular grid of square or cubic blocks with side 2 and computes the max or average over such blocks for each input feature.

While pooling operations are common building blocks of CNNs aiming to reduce the feature map's spatial dimension, it is noteworthy that a similar goal can be achieved using 3×3 convolutions with stride 2 for 2D data. In this case, doubling the number of filters can simultaneously reduce information loss

while aggregating higher-level features. This down-sampling strategy is employed in architectures such as ResNet [191].

3.2.1.3 Fully Connected Layers

A fully connected layer with n input dimensions and m output dimensions is defined as follows. The layer output is determined by the weight matrix $W \in \mathbb{R}^{m \times n}$, which has m rows and n columns, and the bias vector $b \in \mathbb{R}^m$. Given input vector $x \in \mathbb{R}^n$, the output of a fully-connected layer (FC) with activation function f is defined as

$$\mathrm{FC}(x) := f(Wx + b) \in \mathbb{R}^m. \tag{3.43}$$

Here, the function f is applied component-wise. Fully connected layers are typically used as final layers in classification problems, where one or two fully-connected layers are attached on top of a CNN. For this, the CNN output is flattened and treated as a single vector. Another example is various autoencoder architectures, where FC layers are often attached to the latent code in both encoder and decoder paths of the network. When working with CNNs, it is helpful to recognize that for a feature map with n channels, a convolution filter with kernel size 1 and m output channels can be applied, equivalent to applying a fully connected layer with m outputs to each point in the feature map.

3.2.2 Pre-trained CNNs and Transfer Learning

The significant success of CNNs in image recognition has led to the training of extensive networks on vast datasets by large industrial research teams from companies such as Google, Microsoft, and Amazon. Notable models include AlexNet [190], GoogLeNet [221], ResNet [191], InceptionNet [222], and VGGNet [223]. Most researchers and practitioners lack the resources, data, or time to train networks on this scale. Fortunately, pre-trained models have been made available in standard packages like the Torch Vision library in Pytorch [193] or the Caffe framework [224]. These models can be directly used as a basis for fine-tuning in various supervised image recognition tasks through transfer learning.

Transfer learning operates on the principle that the filters (receptive fields) learned by the convolution layers of these networks are informative for most image recognition-based tasks, beyond their initial training purposes. Given that images reflect the natural world, the filters learned by these CNNs are expected to be transferable to new tasks with minor modifications and fine-tuning. This approach has proven effective for numerous tasks. There are three primary methods for repurposing a pre-trained CNN for a new task:

- Using CNN as a *Fixed Feature Detector* at the *Top Layer*. When the new dataset is small and similar to the original dataset, the CNN can be employed as a fixed feature detector while retraining the classifier. This involves removing the classifier (soft-max) layer at the top of the CNN and replacing

it with a new classifier relevant to the supervised learning problem. In this configuration, the CNN acts as a fixed map from images to relevant features.

- Using CNN as a Fixed Feature Detector at an *Intermediate Layer*. For datasets that are small and significantly different from the original training dataset, the features at the top level may not be suitable. In such cases, features from intermediate layers of the CNN can be used to train the new classifier. These intermediate features are generally less fine-tuned and more universal (e.g., edge detectors). This approach is based on the principle that CNNs learn progressively complex features at deeper network layers (refer to the section on representational learning).

- Fine-tuning the CNN. For large datasets, in addition to replacing and retraining the classifier in the top layer, it is possible to fine-tune the weights of the original CNN using back-propagation. During this process, some weights in the CNN may be frozen, or all weights can be retrained simultaneously.

These procedures can be easily executed using packages like Caffe [224] or the Torch Vision library in PyTorch[193].

Chapter 4

CW detection based on a Transfer Learning approach

4.1 Introduction

Continuous gravitational waves (CWs) are a distinct class of Gravitational Wave (GW) signals characterized by the continuous emission of ripples in the fabric of spacetime. These waves are generated by the asymmetric rotation of rapidly spinning neutron stars and pulsars [225]. As these entities spin, their non-spherical configurations induce periodic oscillations in the surrounding spacetime, giving rise to waves possessing a quasi-monochromatic frequency [226]. This frequency remains approximately constant, as long as the rotational dynamics of the object remains stable. Unlike transient GW events originating from compact binary mergers [227–231], which manifest as brief bursts, CWs are a class of GWs that endure over extended periods of time lasting weeks, months or even years [232].

The pursuit of the detection of CWs is primarily driven by the enduring mystery surrounding the pulsar population in our Galaxy. Electromagnetic observations reveal only a few thousand pulsars [233], yet stellar evolution models suggest the existence of over 10^8 compact objects in our vicinity [234]. A significant fraction of these are hypothesized to spin at frequencies between 10 and 1000 Hz, positioning them within the sensitivity band of the second-generation GW detectors. As the sensitivity of these ground-based interferometers improves with each generation [235], the opportunity to detect CWs emitted within our Galaxy becomes increasingly feasible. Nonetheless, the exceedingly low amplitudes of these signals pose substantial challenges [132] for their detection and analysis, rendering them highly susceptible to noisy data [236]. Usually, to increase the signal-to-noise ratio (SNR), it is necessary to integrate as much data as possible. This involves employing *coherent* search methods that integrate signal templates across the entire time-span of the dataset. This approach needs the formulation of a precise model, applying it to the data, and calculating a detection statistic through a huge number of templates. However, the immense computational demand [237] often renders these full-scale *match filtering* searches impractical. Instead, we turn to *semicoherent* search methods that analyze shorter segments coherently and combine

their results in an incoherent way. Given the constraints of a fixed computing budget, these semicoherent methods frequently offer greater sensitivity than their fully-coherent counterparts [238].

In recent years, Deep Learning (DL) techniques have emerged as a powerful alternative for addressing similar complexities [239]. Such approach uses artificial neural networks with multiple layers to automatically learn and extract intricate patterns and representations from complex datasets. In particular, Convolutional Neural Networks (CNNs) are designed for processing grid-like data (such as images or spectrograms) to automatically learn and extract spatially local features, capturing important patterns and structures [240]. One of the key advantages of DL is its ability to learn feature representations from raw data, reducing the need for manual feature engineering and allowing the model to discover complex and non-linear relationships.

In the context of GW physics, CNNs have found diverse applications [241], including parameter estimation of transient signals [242, 243], signal denoising [244], waveform modeling [245, 246], and the detection and classification of glitch noises [247, 248] to name a few. Specifically for CW analysis, DL shows promise in enhancing search sensitivity and efficiency. Dreissigacker and Prix [249] demonstrated the effectiveness of a noise-versus-signal classifier using the *InceptionResNet-v2* architecture [250], which processes strain data from two detectors. Their model performs well over data spans of $T = 10^5$ s compared to traditional match-filtered sensitivities. However, the performance is significantly reduced in the presence of non-Gaussian disturbances, also known as *instrumental lines*. Addressing this, Bayley *et al.* [251] developed a CNN classifier that integrates with the SOAP algorithm [252] to mitigate the impact of these lines present in only one detector. It was shown that this approach increases the robustness of the SOAP searches. Joshi and Prix [253] identified a mismatch between typical state-of-the-art image-classification CNN architectures and the morphology of CW signals. Thus proposing a tailored CNN that preserves the information of input spectrograms, achieving match-filtering sensitivity for targeted CW searches over ten days, though they noted the need for research into longer-duration analyses. Modafferi *et al.* [254] in an innovative approach, explored using CNNs together with partially computed coherent match-filter $\mathcal{F}$-statistic atoms for analyzing GWs from glitching pulsars. This method approaches the efficiency of traditional detection statistics within 10% at a fixed false-alarm probability and shows potential for adaptation to CW searches through minor modifications to the window function.

All of the aforementioned studies have something in common, they developed their Deep Neural Networks (DNNs) from the ground up. In contrast, we propose the use of state-of-the-art, extensively pretrained CNNs, adapting them via *Deep Transfer Learning*. This method involves repurposing the knowledge obtained from different tasks and datasets to reduce learning costs. Typically, an untrained network begins with randomly initialized weights, which are gradually optimized during a costly training process. However, evidence [255] supports that starting with weights from a pretrained network can improve training efficiency over random initialization. This is particularly true for well-established image-recognition networks like *ResNet* and *Inception*, whose deep CNNs, trained on real-world images, possess features directly transferable [256] to the classification of spectrograms obtained from GW time series. For this study, we choose *ResNeXt-50* [257] for three key reasons: its inherent robustness against

overfitting through shortcut connections, superior generalization capabilities relative to predecessors, and a modular architecture that is well suited to accommodate transfer learning tasks.

In this work, we explore the use of transfer learning to develop a binary classifier capable of detecting CW signals immersed in Gaussian and real detector noise with time gaps. We first develop a CNN capable of efficiently detecting CW signals over short observation periods and then evaluate its generalization to spectrograms corresponding to longer observation periods. The network is trained using curriculum learning, starting with the open-source pretrained weights of ResNeXt-50 [257].

This chapter is organized as follows. In Sec. 4.2, we provide a brief overview of the conventional neutron star emission model. Section 4.3 discusses the definitions of our performance measurements. In Sec. 4.4, we present our transfer learning approach and describe the network architecture. Finally, in Sec. 4.5, we characterize the performance of the model and discuss the results for both short observation periods and their generalization to longer periods. Section **??** covers our conclusions and outlines directions for future work.

4.2 Continuous Waves

The strain waveform of CWs originated from an isolated, rotating rigid triaxial ellipsoid (conventional model for a GW-emitting neutron star) can be expressed as [258]

$$\begin{aligned} h(t) = \; & F_+(t,\psi)h_0\frac{1+\cos^2(\iota)}{2}\cos(\Phi(t)) \\ & + F_\times(t,\psi)h_0\cos(\iota)\sin(\Phi(t)), \end{aligned} \tag{4.1}$$

where ι is the angle between the star's spin direction and the propagation direction $\hat{k}$ (pointing towards the Earth), F_+ and $F_\times$ are the detector antenna pattern response factors (See Ref. [259]) to the $+$ and $\times$ polarizations, respectively, which vary with the detector's orientation, the source's location, and the polarization angle ψ. Here, $\Phi(t)$ is the phase of the signal (see Appendix B.1 for more detail) that can be Taylor-expanded as

$$\Phi(\tau) \approx \Phi_0 + 2\pi\sum_{k=0}^{s} f_{\rm ns}^{(k)}\frac{\tau^{k+1}}{(k+1)!}, \tag{4.2}$$

where τ is the proper time in the neutron star rest frame. This expansion models the CWs as long-lasting, quasi-monochromatic waves with a slowly varying instantaneous frequency, that by virtue of Eq. (4.2), can be written as

$$f_{\rm ns} = \frac{1}{2\pi}\frac{d\Phi(\tau)}{d\tau} = \sum_{k=0}^{s} f_{\rm ns}^{(k)}\frac{\tau^k}{k!}. \tag{4.3}$$

While initially one might expect the signal from a rapidly spinning neutron star to appear as a distinct spike at its nominal frequency in the discrete Fourier transform (DFT) of the data, several complexities arise. Given the signal's expected weakness relative to the noise floor, a matched-filtered detection would require a sufficiently long integration time. Hence, for example, if we take $T_{\rm coh} \sim 1$ month, the first derivative of the frequency, $\dot{f}$, would need to satisfy $\dot{f} \lesssim 10^{-12}$ Hz/s to maintain the signal within the

same DFT frequency-bin. However, in practice, Doppler modulation from Earth's motion—typically around 10^{-6} Hz/s—and usually larger $\dot{f}$ values complicate this. For sources with unknown frequency evolution, accounting for these variations significantly increases the computational demands for searching across the parameter space, the so-called *amplitude parameters* $(h_0, \psi, \phi_0, \cos\iota)$ and the *phase-evolution parameters* $(\alpha, \delta, f, \dot{f}, \ddot{f}, \dots)$. To better understand these costs, CW searches are categorized [225] into: i) *Targeted searches* in which the star's position and rotation frequency are known from X-ray and γ-ray observations, rendering the phase-evolution parameters known, ii) *directed searches* for which α and δ are known but the parameters governing the frequency evolution are unknown and iii) *All-sky searches* for unknown neutron stars. The greater the a priori knowledge of source parameters, the more computationally feasible it is to integrate the data coherently for longer time periods.

4.3 Performance Measurements

4.3.1 Benchmark definitions

The sensitivity of a CW search is often described by the upper-limit amplitude $h_0^{p_{\rm det}}$ which is the threshold amplitude where the search attains a specified detection probability $p_{\rm det}$ (commonly set at 90% or 95%) at a false-alarm level $p_{\rm fa}$. This upper-limit amplitude h_0 characterizes a population of signals with unknown amplitude parameters. It represents the maximum amplitude below which signals remain undetectable due to the noise characteristics of the detector, defined by the noise spectral density S_n (a single-sided power spectrum across the frequency band and harmonic mean over time and across multiple detectors). Historically, to better reflect the search-sensitivity independence of the noise level S_n, sensitivity is often expressed by the *sensitivity depth* as

$$\mathcal{D}^{p_{\rm det}} = \frac{\sqrt{S_n}}{h_0^{p_{\rm det}}}. \tag{4.4}$$

Since a smaller $h_0^{p_{\rm det}}$ implies a more sensitive search, $\mathcal{D}$ increases with search sensitivity. By normalizing the sensitivity to the noise power spectrum of the detector, $\mathcal{D}$ quantifies the performance contribution of the CW search algorithm to the overall search sensitivity. It is common to use the sensitivity depth $\mathcal{D}^{90\%}$, corresponding to an upper-limit amplitude $h_0^{90\%}$, for which a search would achieve a detection probability of $p_{\rm det} = 90\%$ at a false-alarm level of 1%. Additionally, the optimal SNR is defined as

$$\rho^2 = \sum_X 4\mathcal{R} \int_0^\infty \frac{\tilde{h}^X(f)\tilde{h}^{X*}(f)}{S_n^X(f)} df, \tag{4.5}$$

which integrates the ratio of the signal's power to the noise spectral density across all frequencies, summing across all detectors X. Analogous to the concept of $\mathcal{D}^{p_{\rm det}}$, we can define the network SNR for which a search would attain a detection probability $p_{\rm det} = 90\%$ at $p_{\rm fa} = 1\%$ as $\mathrm{SNR}^{90\%}$. All of the efficiency curves displayed in this study are based on a false-alarm probability of 1%, which establishes our detection threshold. This threshold is the value of the detection statistic exceeded by only 1% of a large set of noise-only samples. Within the framework of binary-classification, *efficiency* is defined as

the ratio of the number of the relevant retrieved instances to the number of all relevant instances:

$$\text{efficiency} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{4.6}$$

where TP is the number of true positive detections and FN denotes the number of false negatives. This metric essentially measures how effectively a search algorithm can identify signals, minimizing the instances of missed detections.

4.3.2 SOAP CW

The most extensively studied search algorithms, such as the $\mathcal{F}$-statistic method, depend on a predetermined CW model. The parameters of this model are searched and "matched" to an optimal filter that determines whether the data contains a signal. In contrast, other methods do not rely on a predefined model. Instead, they treat the CW signal frequency evolution as a randomly-wandering path across the time-frequency domain. One notable example is the Viterbi algorithm, which is based on a Hidden Markov Model, that searches for the path that yields the highest cumulative fast Fourier transform power [260]. The final result is the most likely signal trajectory through the spectrogram plane.

A variation of this algorithm, known as SOAP, enhances the basic Viterbi approach by incorporating a post-processing stage that uses a CNN to increase robustness against spectral artifacts. This adaptation is intended to be a general-purpose tool optimized for rapid computation. Bayley *et al.* [251] explored various input data formats and network architectures, finding that employing a CNN to analyze the Viterbi map offers a reasonable balance between simplicity and performance. Given the incorporation of a CNN for final predictions, the SOAP CW pipeline provides a valuable benchmark for evaluating the network developed in our study.

4.4 Transfer Learning for Continuous Waves

Transfer learning streamlines the training process by initially pre-training a DNN on a broad, labeled dataset, then fine-tuning it on a more focused dataset of interest. This strategy, particularly prevalent in computer vision tasks [261], leverages the understanding that a CNN's initial layers capture generic features like curves or edges [262]. The final layers, however, learn complex, data-specific characteristics. By reusing a model that has already been trained on a large, diverse dataset and adjusting it for a new, specific one, transfer learning can achieve both higher accuracy and quicker training times.

In this study, we develop two different CNNs: *Gauss ResNeXt*, for which we employ simulated Gaussian noise to create a controlled training environment that allows the network to learn the basic patterns and anomalies associated with these conditions and *O3 ResNeXt*, trained on real detector data, which introduces the complexities and variabilities of actual noise artifacts, such as glitches and non-Gaussian features, providing a more realistic training scenario. In the following sections, we detail the input

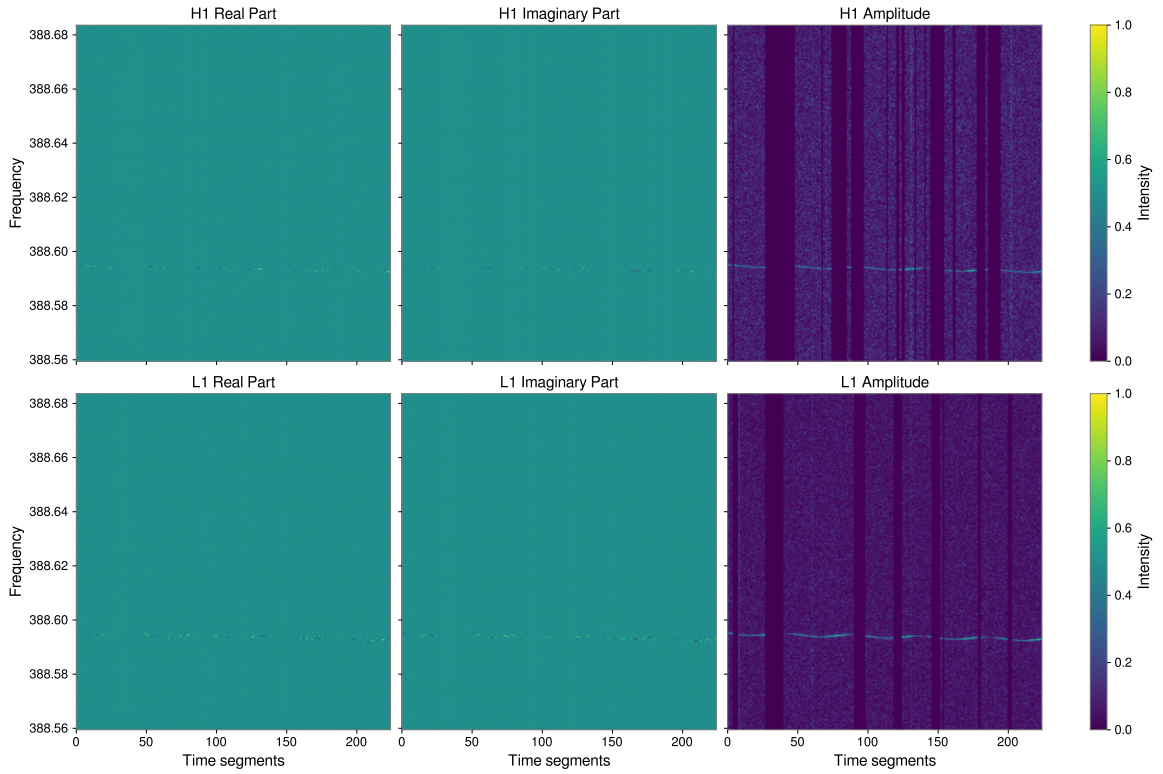


FIGURE 4.1: An example of the input spectrograms used for training the modified ResNeXt CNN. The top panels show the real part, imaginary part, and amplitude of the power spectrograms for data from the H1 detector. The bottom panels show the corresponding spectrograms for the L1 detector. The modified ResNeXt CNN employs a dual-branch architecture, with one branch processing data from the H1 detector and the other processing data from the L1 detector. These spectrograms are stacked along the RGB channels before being fed into the network.

data, architecture, data generation, and training procedures for these models. The input data section discusses the specifics of the spectrograms used, including their dimensions and pre-processing steps. The architecture section provides an in-depth look at the ResNeXt-50 framework and the modifications made to suit our purposes. The data generation section describes how we simulate signals and inject them into both Gaussian and real detector noise. Finally, the training procedures section outlines the curriculum learning strategy, hyperparameter tuning, and optimization techniques employed to fine-tune the models.

4.4.1 Input data

Following the methodologies outlined in Refs. [249, 253], we adopt a time-frequency data representation to accommodate the long duration and well-defined frequency of CWs. This involves segmenting the detector's strain data into 1800-second intervals and performing Fourier transform on each segment. Given that ResNeXt-50 is optimized for 224×224 pixel RGB images, we adapt our spectrogram data accordingly: the real and imaginary parts of the Fourier transform, along with the amplitude, are mapped to the three channels traditionally used for RGB color values. The dimensions of our spectrograms are tailored to the input requirements of ResNeXt-50, with 224 data points in the time axis covering approximately 224×1800 s ~ 4.66 days and 224 points in the frequency axis, equating to a bandwidth

TABLE 4.1: The signal parameters were randomly sampled between these lower and upper limits. Here, we consider one single value for the frequency derivative and a fixed bandwidth size as described in Sec. 4.4.1.

	α [rad]	δ [rad]	f [Hz]	$\cos\iota$	ϕ [rad]	ψ [rad]
Lower limit	0	$-\pi/2$	20	-1	0	0
Upper limit	2π	$\pi/2$	1000	1	2π	$\pi/2$
		$\dot{f} = -10^{-9}$ [Hz/s]			band width = 124 mHz	

of roughly $224 \times (1/1800)$ Hz ~ 0.124 Hz. For the purposes of this study, we refer to this standard spectrogram size as *individual cells*. To address the challenge of instrumental lines, as suggested by Ref. [251], data from both the LIGO Hanford (H1) and LIGO Livingston (L1) detectors are considered. As explained in more detail in Sec. 4.4.2, we employ a binary-classifier network for each detector to distinguish between "signal" and "noise" classes, later integrating the outputs of their respective final layers into a unified *fusion layer* for the conclusive classification task, in turn assembling two networks for a final decision. An example of a typical input spectrogram from both detectors can be found in Fig. 4.1.

4.4.2 Network architecture

In a recent study involving CNNs applied to the detection of CWs using spectrogram data, it has been noted that conventional convolution operations can lead to information loss [253]. This loss is primarily due to downsampling and the transformation of features across layers, which can result in the loss of critical fine-grained details. This issue is well-documented in computer vision, inspiring the development of *residual connections* [263, 264]. Also known as shortcut connections, they perform identity mapping by directly passing the input of a layer to its output. This mechanism ensures that the original information is preserved and available at deeper layers, hence addressing the potential loss caused by convolutional operations. By allowing earlier features to be reused in later layers, residual connections ensure that essential information, which might be diluted or transformed unfavorably by successive convolutions, is retained and can contribute to the final representation. This feature is a key component of the ResNeXt architecture and is the main reason for selecting this particular network for our study.

For our analysis, we developed a two-stage network architecture. It comprises 1) two adapted ResNeXt-50 models processing H1 and L1 detector spectrograms, respectively, and 2) a fusion layer that integrates the classifiers' outputs into a unified final prediction, serving as a detection statistic. The structure of the network is shown in Fig. 4.2. The networks were implemented in PyTorch 2.0 [265], which incorporates a new model compilation feature for faster training and inference. The training was conducted on a 16 GB NVIDIA Tesla V100 GPU.

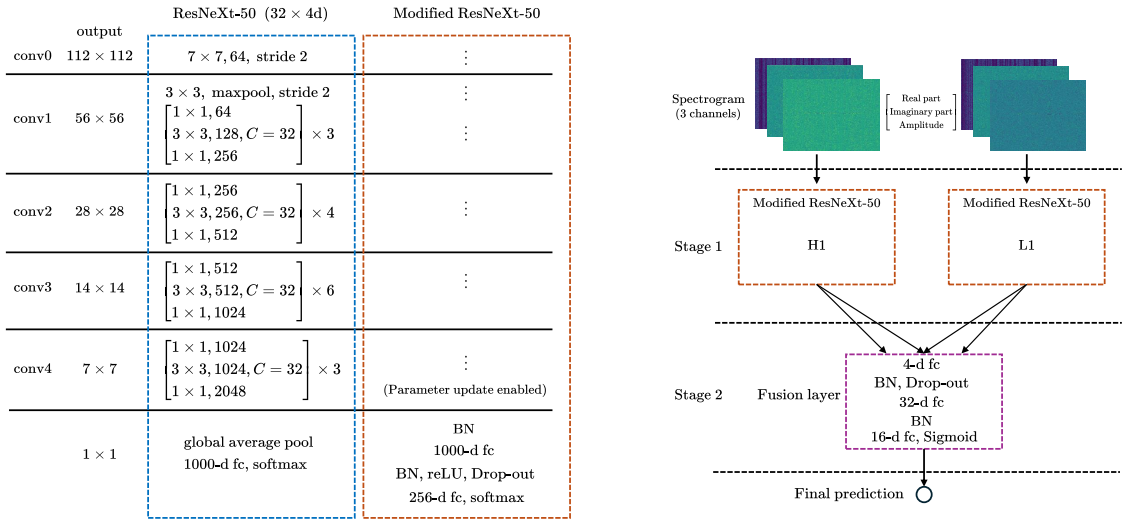


FIGURE 4.2: (**Left**) ResNeXt-50 with a 32 × 4d architecture and our modified version. Inside the brackets we find the shape of a residual block, and the value outside the brackets represent the number of stacked blocks. $C = 32$ denotes grouped convolutions with 32 groups. The main difference between the original and the modified network lies in the final fully-connected layer that has been replaced for a customized one. The parameters of the 4th convolutional layer are allowed to be updated during training while the rest of the parameters remain frozen. (**Right**) Structure of the ensemble network. (Stage 1) The modified ResNeXt-50 units take the spectrograms from the H1 and L1 detectors as inputs. Their outputs are then combined in a *Fusion layer* (Stage 2) that gives the final prediction of the entire combined network.

4.4.2.1 Modified ResNeXt-50

The initial stage makes use of a custom DL classifier based on the `resnext50_32×4d` architecture. It initializes with the pretrained open-source weights learned from the *ImageNet-1K* dataset. To adjust the network to our specific classification task, we enable parameter updates for the fourth convolutional layer during training to refine complex feature representations. Additionally, we replace the original fully-connected output layer with a customized sequence of fully-connected layers that incorporate batch normalization and dropout for regularization, ultimately producing two outputs for class prediction.

4.4.2.2 Fusion layer

To reduce the impact of instrumental lines, predictions from both H1 and L1-trained networks are merged. The fusion layer first concatenates these inputs, then processes them through two dense layers equipped with batch normalization and dropout, culminating in a final classification layer. This integration should be able to introduce a level of redundancy and cross-validation between both data sources, making the prediction more robust and accurate.

4.4.3 Data generation and preprocessing

This section elaborates on the methodology employed to generate training data for evaluating the performance of networks trained solely on Gaussian noise versus those trained exclusively on real-detector

data. A special emphasis is made in the incorporation of realistic time gaps to mimic actual operational conditions. For Gaussian noise simulations, we analyzed the Data Quality flags of the `O3a_16KHZ_R1` channel taken from data spanning observation run O3a to construct an empirical probability distribution of time-gap occurrences and durations. Conversely, for real detector data which inherently contains time gaps, we utilized three months worth of observations from the O3 run. Then, we randomly extract 20 slices each spanning 4.66 days as the basis for our noise Short-time Fourier Transform (SFT) files covering the frequency band from 20 to 1000 Hz. All the real detector data was retrieved from the Gravitational Wave Open Science Center [266].

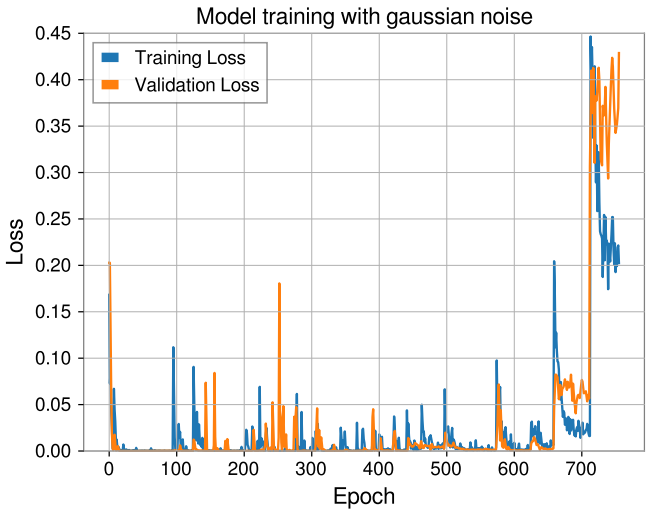
Simulated signals were generated using the PyFstat package [267, 268], with their parameters randomly selected from predefined prior distributions (referenced in Table 4.1). The SNR calculations were performed using the `SignalToNoiseRatio.from_sfts` function within PyFstat, that uses the definition of optimal SNR as described in Ref. [269]. Subsequently, the amplitude h_0 of the CW is scaled accordingly to achieve the desired SNR across the detector network. The PyFstat package takes care of the generation of the right antenna pattern functions and the distribution of the power spectrum of the signal.

For effective training, a substantial volume of labeled data, categorized as either "signal" or "noise," is necessary. To add diversity to the dataset, each sample comprises various augmentations, namely random rolling of spectrograms along the frequency axis and flipping along the time axis. Each network undergoes training with 40,000 samples, evenly split between software-injected signals and pure noise. As detailed in Sec. 4.4.4, the total number of samples are divided into 20 batches, each containing increasing SNRs from 10 to 200.

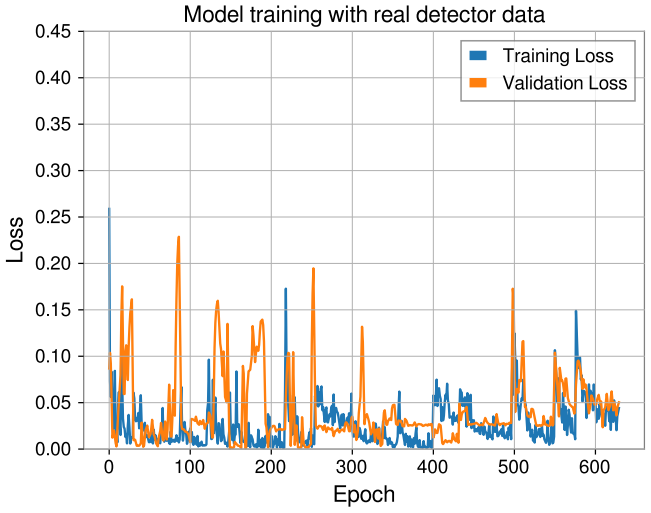
4.4.4 Network training and validation

In this study, we employed a curriculum learning strategy [270, 271], taking the SNR as a measure of difficulty, with values ranging from 30 to 200 in increments of 10. This method begins training with "easy" high-SNR signals, using an implementation of an early-stopping mechanism to retain weights yielding the lowest validation loss observed across the preceding 20 epochs. As the network progresses to increasingly challenging signal batches, the fine-tuning of hyperparameters—including the initial learning rate and dropout rates for the final dense layers in both the modified ResNeXt-50 and fusion layer—is conducted. This optimization relies on the Optuna framework [272] to perform a grid search across 50 distinct hyperparameter combinations. Additionally, L2 regularization was applied, featuring a weight decay factor of 10^{-5}. On average, the network was trained for approximately 40 epochs per batch. During the training phase of our neural network models, we employed the Adam optimizer [273] for its efficient computation and use of adaptive learning rates. For the loss function, cross-entropy loss [274] was chosen due to its suitability for binary and multi-class classification tasks. It provides a direct measure of the discrepancy between the predicted probabilities and the actual labels.

The training processes of Gauss ResNeXt and O3 ResNeXt is shown in Fig. 4.3. During the last phase of training, distinct patterns emerged for models trained with Gaussian noise and real detector noise. For Gauss ResNeXt, there was a significant increase in both training and validation losses, possibly indicating



Training of the Gauss ResNeXt network.



Training of the O3 ResNeXt network.

FIGURE 4.3: Training and validation loss of the trained networks.

model capacity limits when exposed to the most challenging examples. Conversely, the training of O3 ResNeXt is marked by considerable fluctuations throughout the process, which reflects the complexity and variability of the real noise environment. Although persistent fluctuations suggested ongoing difficulties in adapting during the late phase, the overall convergence of both patterns indicates that the model has avoided overfitting and achieved a more stable state. The models' complete training process takes around 20.3 hours to be completed.

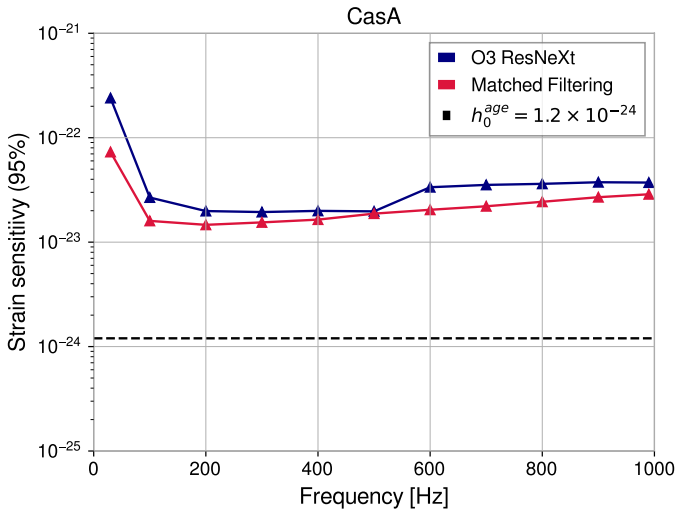


FIGURE 4.4: Targeted search for the Cassiopeia A supernova remnant. O3 ResNeXt performance is similar to the Matched filtered one provided by the `WEAVE` algorithm.

4.5 Results and Discussion

4.5.1 Detection efficiency for individual cells

4.5.1.1 Performance check with a targeted search for the Cassiopeia A

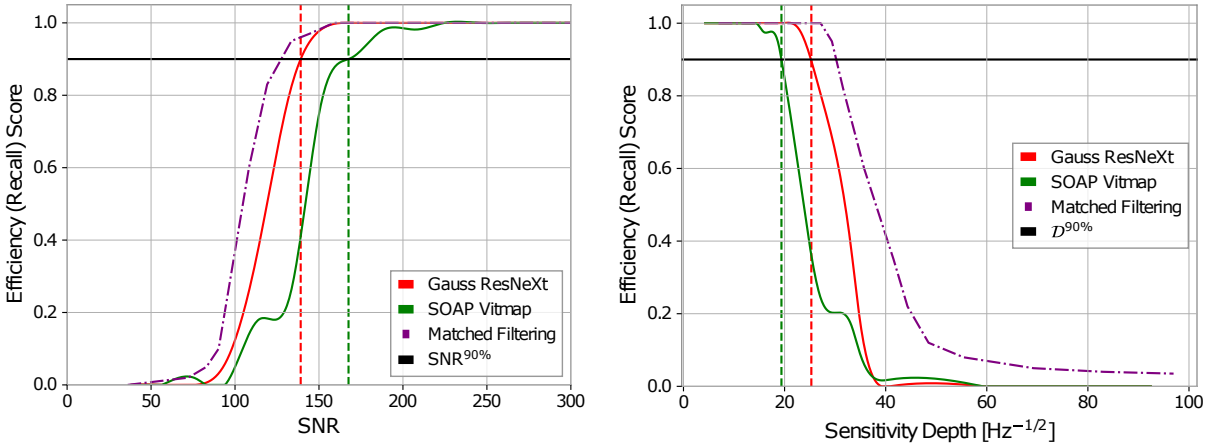
As is customary for machine learning applications, to characterize the performance of our CW detection method, it is necessary to evaluate it on independent test sets. The first test involves a targeted search for the supernova remnant Cassiopeia A [275] with an age-based limit of $\sim 1.2 \times 10^{-24}$, using real detector noise and signals with randomly selected phase parameters. Specifically, we aim to determine the GW strain required for the detection method to achieve a 90% detection efficiency across frequencies ranging from 20 to 1000 Hz. For comparison, we conducted a `WEAVE` search [276], which is based on a semi-coherent averaging of $\mathcal{F}$-statistic values over many individual intervals, each with a coherence time $T_{\text{coh}} = 30$ minutes. In Gaussian noise without a signal, the value of $2\mathcal{F}$ follows a χ^2 distribution with four degrees of freedom and an expectation value of four. The presence of signals, however, results in a non-central χ^2 distribution with a non-centrality parameter proportional to h_0^2 and depends on the sky location, source orientation, and the detectors' orientations and locations. This offset provides an estimate for the matched-filtering detection statistic. As shown in Fig. 4.4, the O3 ResNeXt demonstrates competitive performance compared to the matched-filtering alternative, with only a slight decrease in performance at frequencies higher than 500 Hz. This decrease is commonly attributed to the increasing Doppler broadening of the signal in the detector frame, which is a well-known effect.

4.5.1.2 Efficiency versus signal strength

To evaluate the models, we generated two independent test datasets: one containing Gaussian noise data and the other containing real detector data, with signals of varying depths $\mathcal{D}$ from 10 to 150,

corresponding to SNRs from 5 to 400. For the real detector noise, we selected ten new data chunks different from those used in training. We measured the efficiency of Gauss ResNeXt and O3 ResNeXt models at a fixed false-alarm rate of 1%. For comparison, we included the performance of the established SOAP Viterbi map (Vitmap) method, as illustrated in Figs. 4.5 and 4.6.

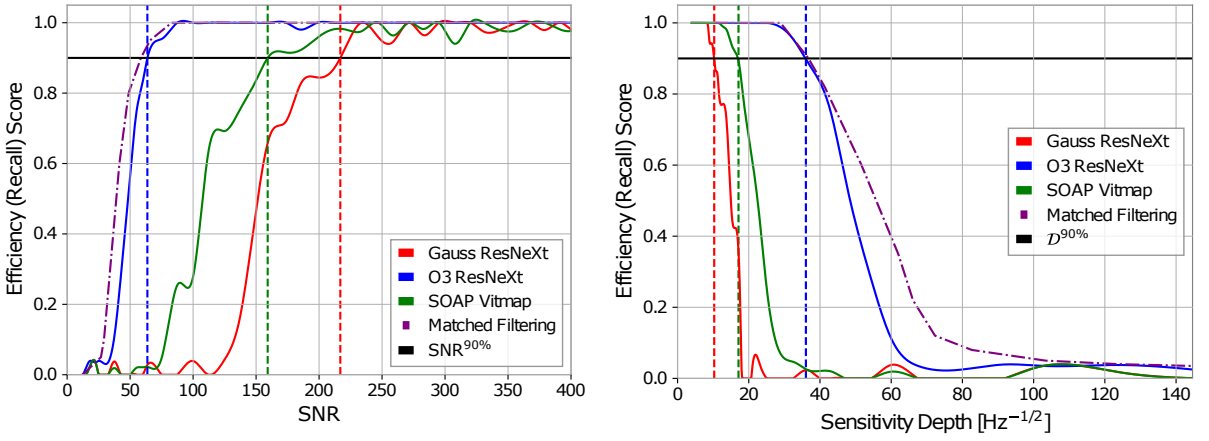
In the first scenario, Fig. 4.5 depicts the models' performance on signals injected into Gaussian noise. Both Gauss ResNeXt and SOAP Vitmap exhibited similar sensitivities. However, O3 ResNeXt performed poorly on this dataset, misclassifying all samples as "signal." In the second scenario, Fig. 4.6 shows the efficiency of the models on signals injected into real noise data. Here, O3 ResNeXt outperformed the other models, followed by SOAP Vitmap and Gauss ResNeXt. This disparity of the generalization capability of O3 ResNeXt reveals a key observation: the network trained on Gaussian noise data performs poorly



(a) Efficiency vs. SNR for signals injected into Gaussian noise.

(b) Efficiency vs. Sensitivity Depth for signals injected into Gaussian noise.

FIGURE 4.5: Efficiency of search for signals injected into Gaussian noise. The efficiency is computed with a threshold set by the 1% false alarm rate. The most efficient model is Gauss ResNeXt, which was trained on Gaussian noise data, followed by the SOAP Vitmap. The O3 ResNeXt (trained with real detector noise) failed to produce any significant result and was largely misclassifying every sample as signal.



(a) Efficiency vs. SNR for signals injected into real detector noise.

(b) Efficiency vs. Sensitivity Depth for signals injected into real detector noise.

FIGURE 4.6: Efficiency of search for signals injected into real detector noise from the O3 observation run. The most efficient model is the one trained with real detector data, followed by the SOAP CW Vitmap and Gauss ResNeXt coming at last.

TABLE 4.2: Network detection efficiency at fixed 1% false-alarm rate attained by the trained networks and the SOAP CW Vitmap evaluated on Gaussian/Real detector noise datasets.

	$\mathcal{D}^{90\%}$		$\mathrm{SNR}^{90\%}$	
	Gaussian noise	Real noise	Gaussian noise	Real noise
Matched-Filtering	30.83	36.78	127.54	59.72
Gauss ResNeXt	25.29	9.92	139.09	215.08
O3 ResNeXt	−−	36.04	−−	63.16
SOAP Vitmap	19.37	17.11	167.57	157.90

on real noise data, and vice versa. This suggests that each model learns noise characteristics specific to its training noise environment. While O3 ResNeXt excels with real detector data, its training on complex, realistic noise environments likely gives it the ability to detect weaker signals more effectively, but at the cost of performance in simpler, Gaussian noise scenarios. This apparent overfitting to their noise environment is caused by the very deep architecture of the base ResNeXt-50 model, that learns deep features of the underlying spectrogram background. The SNRs and signal depths $\mathcal{D}$ at which the models achieved 90% efficiency are summarized in Table 4.2. For both Gauss ResNeXt and O3 ResNeXt, we observe performance close to matched-filtering in their corresponding noise environments. However, these good results deteriorate at sensitivity depths > 25 Hz$^{-1/2}$ for Gauss ResNeXt and > 40 Hz$^{-1/2}$ for O3 ResNeXt. Despite this, we consider the performance of O3 ResNeXt to be sufficiently robust to attempt generalization to larger spectrogram sizes.

4.5.2 Generalization to larger cells

Next, we turn our attention to longer observation times and potentially wider frequency bands. There are two perspectives to consider: either retraining the model with entire spectrograms encompassing larger observation periods or splitting a long-duration spectrogram into smaller segments, calculating a

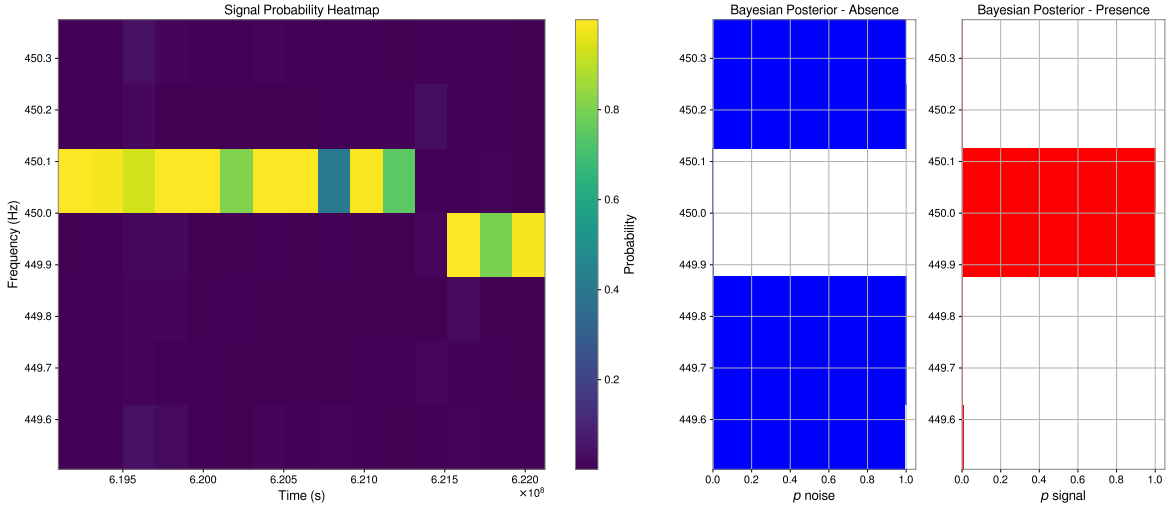


FIGURE 4.7: Example of a larger spectrogram composed of 7×15 individual cells, encompassing a bandwidth of 0.868 Hz and a duration of 69.9 days. The left panel shows the signal probability heatmap, indicating the probability of signal presence across time and frequency. The right panels display the Bayesian posterior distributions for noise (absence) and signal (presence) as a function of frequency.

detection statistic for each, and combining them later. This brings us back to the well-known debate of coherent versus semicoherent methods. Due to its computational feasibility and the fact that it requires no additional training, we chose the latter approach. Our goal is to determine if the method used for individual cells can be generalized by using a semicoherent framework to combine the detection statistics of multiple individual cells. For simplicity, we employ a naive Bayesian approach to calculate the posterior probability of noise/signal for the i-th frequency bin as

$$p_i(C_k|\boldsymbol{x}) = \frac{1}{Z} p(C_k) \prod_{j=1}^{n} p(x_j|C_k), \tag{4.7}$$

where $Z = \sum_k p(C_k)p(\boldsymbol{x}|C_k)$ is a normalization factor. The posterior p_i represents a probability for each of the $k = 0, 1$ possible classes $C_k = \{\text{noise}, \text{signal}\}$ given a problem instance to be classified, represented by a vector $\boldsymbol{x} = (x_1, \ldots, x_n)$ encoding n individual cells along the time axis. By calculating this over each frequency bin, we obtain a posterior distribution as a function of frequency. Then to assign a class label $\hat{y} = C_k$ to the i-th frequency bin, we apply the *maximum a posteriori* decision rule, which is defined as

$$\hat{y}_i = \underset{k\in\{0,1\}}{\arg\max}\ p(C_k) \prod_{j=1}^{n} p(x_j|C_k). \tag{4.8}$$

Finally, if for a given large spectrogram, any of the $\hat{y}_i$ is classified as "signal" rather than "noise," the final classification for the entire spectrogram is "signal". A schematic on the segmentation of larger spectrograms into individual cells is shown in Fig. 4.7.

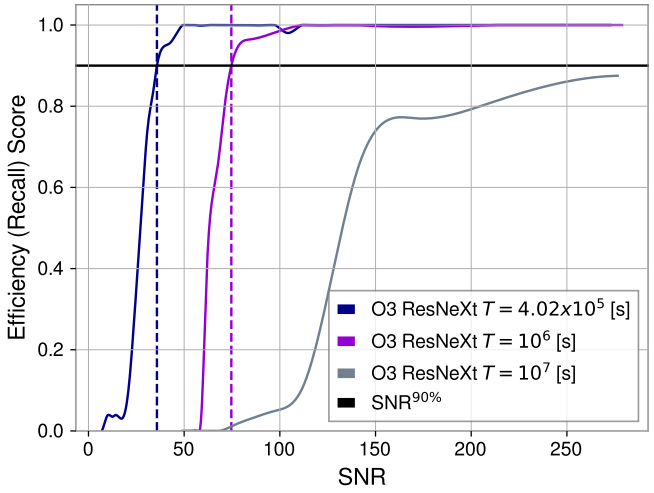
To test this method, we use the O3 ResNeXt network to evaluate datasets containing two distinct sizes of spectrograms with signals embedded in real-detector noise. Both datasets cover a frequency bandwidth of 0.868 Hz, corresponding to seven individual cells along the frequency axis. The time durations for these datasets are: a) five individual cells along the time axis, covering $\sim 10^6$ seconds, and b) 25 individual cells along the time axis, covering $\sim 10^7$ seconds. As shown in Fig. 4.8, the efficiency decreases when dealing with longer duration spectrograms. Table 4.3 summarizes the signal strength at 90% efficiency.

Instrumental artifacts likely play a significant role in the observed decrease in performance for the spectrograms considered in this study. As noted in Ref. [277], the O3 strain data from the detectors contains many contaminated sub-bands that were removed from their analysis due to saturation and severe contamination. However, in our study, we chose not to exclude these sub-bands to assess the robustness of our model against such disturbances.

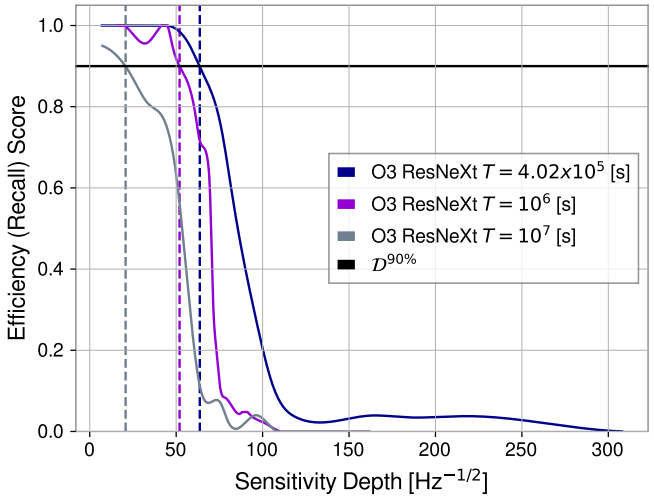
Given the limitations discussed earlier, there are several alternatives to enhance the work presented here. The successful detection capabilities observed for $T \sim 4.66$ days should, in principle, be extendable

TABLE 4.3: Detection efficiency (1% false alarm rate) of the O3 ResNeXt for larger duration spectrograms.

	$\mathcal{D}^{90\%}$	$\text{SNR}^{90\%}$
O3 ResNeXt $T = 4.02 \times 10^5$ s	63.16	36.01
O3 ResNeXt $T = 10^6$ s	51.59	75.28
O3 ResNeXt $T = 10^7$ s	20.60	$--$



Efficiency versus SNR for signals injected into real detector noise



Efficiency versus sensitivity depth for signals injected into real detector noise

FIGURE 4.8: Efficiency of the O3 ResNeXt model when evaluating larger spectrograms containing signals injected into real detector noise. As the observation time increases, we see a clear performance deterioration.

to longer durations. First, to address the issue of instrumental artifacts, an additional classifier could be integrated into the proposed architecture, classifying into three categories: 1) signal, 2) noise, and 3) artifact. This should enhance robustness and add statistical stability when generalizing to longer spectrograms. Another possibility, as discussed by Ref. [277], is to veto frequency bins excessively contaminated with non-stationary noise and spikes. However, ideally, we would prefer an automatic tool that does not require a manually selected list of frequency bins.

Another problem is how to effectively combine the individual detection statistics for each cell, as formulated in Sec. 4.5.1.2. The naive Bayes classifier is a preliminary approach. However, as noted by Chandra [278], updating the posterior probability (Eq. (4.7)) based on the probability of signal presence/absence can lead to *numerical underflow* and floating-point errors due to the extremely small magnitudes involved. Therefore, a more stable approach is to use log-probabilities instead of direct probabilities.

Instead of computing the posterior using Eq. (4.7), we compute a version using log-probabilities for numerical stability. For each frequency bin i, the posterior probabilities are calculated as follows

$$\log p_i(C_k|\boldsymbol{x}) = \log p(C_k) + \sum_{j=1}^{n} \log p(x_j|C_k) \tag{4.9}$$

which, upon normalization and conversion back to probabilities, becomes

$$p_i(C_k|\boldsymbol{x}) = \frac{\exp(\log p_i(C_k|\boldsymbol{x}))}{\sum_k \exp(\log p_i(C_k|\boldsymbol{x}))} \tag{4.10}$$

Another alternative is to use the *logits*, which are the raw outputs of the modified ResNeXt before the final sigmoid layer, as the detection statistic. In other words, we would have a statistic $\hat{\theta} \in \mathbb{R}$, similar to the $\mathcal{F}$-statistic, making possible the use of an analogous approach and enabling combination across the time axis via the same theoretical and programming framework.

Finally, an alternative method involves taking the statistics obtained for the entire frequency bin and combining them into a vector that can be fed to another Multi-layer Neural Network for the final prediction. Any of the methods and improvements mentioned before, if performed well, should approximate the optimal filter. As demonstrated by Joshi and Prix [253], the statistic output by a CNN approaches the functional form of the optimal semi-coherent $\mathcal{F}$-statistic up to an affine transformation. This phenomenon should be observed here when properly combining the detection scores.

4.6 Conclusions

In this study, we developed a binary classifier CNN using transfer learning to detect unmodeled CW signals in Gaussian and real detector noise. Utilizing the pre-trained ResNeXt-50 architecture as our foundation, we trained two models, Gauss ResNeXt and O3 ResNeXt, employing a curriculum learning strategy. Compared to the SOAP CW Vitmap method, our transfer learning approach demonstrated superior performance. For short time durations of approximately 4.66 days, the models achieved sensitivities close to those of matched-filtering. However, both networks learned features specific to their respective noise environments and generalized poorly when evaluated on different types of noise.

To extend observation periods and frequency bands, we adopted a Naive Bayes classifier approach, using the previously developed network to segment large spectrograms into individual cells and combine the detection statistics for each frequency bin. The performance of O3 ResNeXt decreased with longer observation periods, revealing challenges in maintaining performance over extended durations. Therefore, further work is required to improve detection efficiency. Exploring alternative and more robust approaches than the Naive Bayes classifier to combine the detection statistics of individual cells could enhance performance in larger spectrograms. Another potential improvement is to utilize this network architecture for parameter estimation given a prior physical model.

One significant challenge we encountered was the numerical instability associated with combining probabilities directly, possibly due to underflow issues. To address this, the use of log-probabilities for computing posterior distributions should be introduced, which should provide a more stable numerical framework. This approach involves calculating the log-probabilities for signal presence and absence, then normalizing these to obtain the final posterior probabilities. This method can potentially the numerical issues encountered with the traditional Naive Bayes classifier.

Another promising approach involves using the raw logits from the network before the final sigmoid layer as detection statistics. This method allows for a broader range of values, similar to the $\mathcal{F}$-statistic, facilitating the combination of detection scores across the time axis more effectively. Additionally, integrating a classifier to differentiate between signal, noise, and instrumental artifacts could enhance the model's robustness and statistical stability.

Finally, combining the detection statistics from all frequency bins into a vector that can be fed into another Multi-layer Neural Network for final prediction presents another potential improvement. Any of the methods mentioned before, if properly optimized, theoretically, could approximate the optimal filter. As demonstrated by Joshi and Prix, the statistic output by a CNN approaches the functional form of the optimal semi-coherent $\mathcal{F}$-statistic up to an affine transformation, suggesting similar improvements could be achieved here with proper training.

Future work should focus on refining the proposed architecture to enhance its robustness against instrumental artifacts. An intriguing avenue for future research is expanding the current network for parameter estimation of CW signals. Initially, this could involve estimating the parameters of an isolated neutron star. Subsequently, the model could be generalized to a binary system, incorporating additional orbital parameters that add complexity to the signal characterization.

Appendix A

A.1 Einstein Field Equations

The Einstein Field Equations can be derived from the principle of least action. This involves integrating a Lagrangian density over space-time:

$$S = \int d^n x \, \mathcal{L}. \tag{A.1}$$

The Lagrangian density is a scalar, and the only independent scalar that can be constructed from the metric, which includes second derivatives, is the Ricci scalar R. Hilbert demonstrated that the simplest choice for such a Lagrangian is

$$\mathcal{L}_H = \sqrt{-g}\, R. \tag{A.2}$$

The equations of motion are obtained by varying the action with respect to the metric. Using the identity $R = g^{\mu\nu} R_{\mu\nu}$, we get

$$\begin{aligned} \delta S_H &= \int d^n x \, \left(\sqrt{-g} g^{\mu\nu} \delta R_{\mu\nu} + \sqrt{-g} R_{\mu\nu} \delta g^{\mu\nu} + R \, \delta \sqrt{-g} \right) \\ &= (\delta S)_1 + (\delta S)_2 + (\delta S)_3. \end{aligned} \tag{A.3}$$

The second term, $(\delta S)_2$, is explicitly expressed as a factor multiplied by $\delta g^{\mu\nu}$. For the first term, it can be shown that

$$\delta R^{\mu}{}_{\nu\rho\sigma} = \nabla_\rho \left(\delta \Gamma^{\mu}{}_{\sigma\nu} \right) - \nabla_\sigma \left(\delta \Gamma^{\mu}{}_{\rho\sigma} \right), \tag{A.4}$$

where $\Gamma^{i}{}_{kl}$ are the Christoffel symbols given by

$$\Gamma^{i}{}_{kl} = \frac{1}{2} \left(\frac{\partial g_{mk}}{\partial x^l} + \frac{\partial g_{ml}}{\partial x^k} - \frac{\partial g_{kl}}{\partial x^m} \right). \tag{A.5}$$

Hence, the contribution of the first term is

$$(\delta S)_1 = \int d^n x \, \sqrt{-g} g^{\mu\nu} \left(\nabla_\rho \left(\delta \Gamma^{\rho}{}_{\nu\mu} \right) - \nabla_\nu \left(\Gamma^{\rho}{}_{\rho\mu} \right) \right) \tag{A.6}$$

$$= \int d^n x \, \sqrt{-g} \nabla_\sigma \left(g^{\mu\sigma} \left(\delta \Gamma^{\rho}{}_{\nu\mu} \right) - g^{\mu\nu} \left(\delta \Gamma^{\sigma}{}_{\mu\nu} \right) \right). \tag{A.7}$$

The integrand in the previous expression is a total derivative that, by virtue of Stokes' theorem, contributes only with a boundary term. When the variation of the metric $\delta g^{\mu\nu}$ cancels out in the vicinity of the boundary, $(\delta S)_1$ equals zero. For the third term, $(\delta S)_3$, we use the following identity valid for any matrix M:

$$\mathrm{Tr}(\ln M) = \ln(\det M). \tag{A.8}$$

The variation of this identity implies

$$\mathrm{Tr}(M^{-1}\delta M) = \frac{1}{\det M}\delta(\det M), \tag{A.9}$$

then we have

$$\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}g_{\mu\nu}\delta g^{\mu\nu}. \tag{A.10}$$

Replacing this in Eq. (A.3) and considering that the first term has a null contribution:

$$\delta S_H = \int d^n x\sqrt{-g}\left\{R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}\right\}\delta g^{\mu\nu}. \tag{A.11}$$

This integral must vanish for arbitrary variations, leading to the Einstein Field Equation in vacuum:

$$\frac{1}{\sqrt{-g}}\frac{\delta S}{\delta g^{\mu\nu}} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 0. \tag{A.12}$$

To obtain the full EFE, we need to consider an action of the form

$$S = \frac{1}{16\pi}S_H + S_M, \tag{A.13}$$

where S_M is the action for matter. Analogous to the previous derivation, we find

$$\frac{1}{\sqrt{-g}}\frac{\delta S}{\delta g^{\mu\nu}} = -\frac{1}{16\pi}\left(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}\right) + \frac{1}{\sqrt{-g}}\frac{\delta S_M}{\delta g^{\mu\nu}} = 0, \tag{A.14}$$

from which the EFE are recovered if

$$T_{\mu\nu} = \frac{2}{\sqrt{-g}}\frac{\delta S_M}{\delta g^{\mu\nu}}. \tag{A.15}$$

A.2 Wave equation derivation

The starting point is the expansion of the EFE around the flat-space metric

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad |h_{\mu\nu}| \ll 1, \tag{A.16}$$

expanding the Christoffel symbols up to linear oder in $h_{\mu\nu}$

$$\begin{aligned}\Gamma^{\rho}{}_{\mu\nu} &= \frac{1}{2}g^{\rho\sigma}\left(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}\right) \\ &= \frac{1}{2}\left(\eta^{\rho\sigma} - h^{\rho\sigma}\right)\left(\partial_\mu\left[\eta_{\sigma\nu} + h_{\sigma\nu}\right] + \partial_\nu\left[\eta_{\sigma\nu} + h_{\sigma\nu}\right] - \partial_\sigma\left[\eta_{\sigma\nu} + h_{\sigma\nu}\right]\right) \\ &= \frac{1}{2}\left(\partial_\mu h_{\sigma\nu} + \partial_\nu h_{\sigma\mu} - \partial_\sigma h_{\mu\nu}\right) + \mathcal{O}(h\partial h)\end{aligned} \tag{A.17}$$

For the Riemann tensor we have

$$\begin{aligned}R^{\mu}{}_{\nu\rho\sigma} &= \partial_\sigma\Gamma^{\mu}{}_{\nu\sigma} - \partial_\sigma\Gamma^{\mu}{}_{\nu\rho} + \Gamma^{\mu}{}_{\alpha\rho}\Gamma^{\alpha}{}_{\nu\sigma} - \Gamma^{\mu}{}_{\alpha\sigma}\Gamma^{\alpha}{}_{\nu\rho} \\ &= \frac{1}{2}\Big\{\partial_\rho\eta^{\mu\alpha}\left(\partial_\nu h_{\alpha\sigma} + \partial_\sigma h_{\alpha\nu} - \partial_\alpha h_{\nu\sigma}\right) - \partial_\sigma\eta^{\mu\alpha}\left(\partial_\nu h_{\alpha\rho} + \partial_\rho h_{\alpha\nu} - \partial_\alpha h_{\nu\rho}\right)\Big\} + \mathcal{O}\left((\partial h)^2\right)\end{aligned} \tag{A.18}$$

so the Ricci tensor is up to first order

$$\begin{aligned}R_{\nu\sigma} &= R^{\beta}{}_{\nu\beta\sigma} \\ &= \frac{1}{2}\left\{\partial_\beta\eta^{\beta\alpha}\left(\partial_\nu h_{\alpha\sigma} + \partial_\sigma h_{\alpha\nu} - \partial_\alpha h_{\nu\sigma}\right) - \partial_\sigma\eta^{\beta\alpha}\left(\partial_\nu h_{\alpha\beta} + \partial_\rho h_{\alpha\nu} - \partial_\alpha h_{\nu\beta}\right)\right\} \\ &= \frac{1}{2}\eta^{\beta\alpha}\left\{\partial_\beta\partial_\nu h_{\alpha\sigma} + \partial_\beta\partial_\sigma h_{\alpha\nu} - \partial_\beta\partial_\alpha h_{\nu\sigma} - \partial_\sigma\partial_\nu h_{\alpha\beta} - \partial_\sigma\partial_\beta h_{\alpha\nu} + \partial_\sigma\partial_\alpha h_{\nu\beta}\right\} \\ &= \frac{1}{2}\eta^{\beta\alpha}\underbrace{\left\{\partial_\beta\partial_\nu h_{\alpha\sigma} - \partial_\beta\partial_\alpha h_{\nu\sigma} + \partial_\sigma\partial_\alpha h_{\nu\beta} - \partial_\sigma\partial_\nu h_{\alpha\beta}\right\}}_{C_{\beta\nu\alpha\sigma}}\end{aligned} \tag{A.19}$$

Then, the Ricci scalar can be described as

$$R = R^{\sigma}{}_{\sigma} = \eta^{\sigma\nu}R_{\nu\sigma} = \frac{1}{2}\eta^{\beta\alpha}\eta^{\sigma\nu}C_{\beta\nu\alpha\sigma} \tag{A.20}$$

Replacing Eqs. (A.18), (A.19) and (A.20) into the left-hand side of the EFE yields

$$\begin{aligned}G_{\mu\nu} &= R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} \\ &= \frac{1}{2}\eta^{\beta\alpha}C_{\beta\mu\alpha\nu} - \frac{1}{2}\eta_{\mu\nu}\left(\frac{1}{2}\eta^{\beta\alpha}\eta^{\sigma\rho}C_{\beta\rho\alpha\sigma}\right) \\ &= \frac{1}{4}\left(2C^{\alpha}{}_{\mu\alpha\nu} - \eta_{\mu\nu}C^{\rho\alpha}{}_{\rho\alpha}\right) \\ &= \frac{1}{4}\Big(2\partial^\alpha\partial_\mu h_{\alpha\nu} - 2\Box h_{\mu\nu} + 2\partial_\nu\partial_\alpha h_{\mu}{}^{\alpha} - \partial_\nu\partial_\mu h - \eta^{\mu\nu}\left(2\partial^\alpha\partial_\rho h_{\alpha}{}^{\rho} - 2\Box h\right)\Big) \\ &= \frac{1}{2}\Big(\partial^\alpha\partial_\mu h_{\alpha\nu} + \partial^\alpha\partial_\nu h_{\mu\alpha} - \Box h_{\mu\nu} + \partial_\nu\partial_\mu h - \eta^{\mu\nu}\partial^\alpha\partial_\rho h_{\alpha}{}^{\rho} + \eta^{\mu\nu}\Box h\Big)\end{aligned} \tag{A.21}$$

Where $h = \eta^{\mu\nu}h_{\mu\nu}$ and $\Box = \partial^\mu\partial_\alpha$. The equations of motion are written more compactly defining

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h, \tag{A.22}$$

which implies

$$\bar{h} = \eta^{\mu\nu}\bar{h}_{\mu\nu} = \eta^{\mu\nu}\left(h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h\right) = h - \frac{1}{2}(4) = -h \tag{A.23}$$

Hence

$$h_{\mu\nu} = \bar{h}_{\mu\nu} - \frac{1}{2}\eta^{\mu\nu}\bar{h} \tag{A.24}$$

Plugging this result into Eq. (A.21) we get

$$\begin{aligned}
G_{\mu\nu} &= \frac{1}{2}\Big\{\partial^\alpha\partial_\mu\left(\bar{h}_{\alpha\nu} - \frac{1}{2}\eta^{\alpha\nu}\bar{h}\right) + \partial^\alpha\partial_\nu\left(\bar{h}_{\alpha\nu} - \frac{1}{2}\eta^{\alpha\nu}\bar{h}\right) - \Box\left(\bar{h}_{\alpha\nu} - \frac{1}{2}\eta^{\alpha\nu}\bar{h}\right) \\
&\quad - \partial_\nu\partial_\mu(-\bar{h}) - \eta_{\mu\nu}\partial^\alpha\partial^\rho\left(\bar{h}_{\alpha\nu} - \frac{1}{2}\eta^{\alpha\nu}\bar{h}\right) + \eta_{\mu\nu}\Box(-\bar{h})\Big\} \\
&= \frac{1}{2}\Big\{\partial^\alpha\partial_\mu\bar{h}_{\alpha\nu} + \partial^\alpha\partial_\nu\bar{h}_{\mu\alpha} - \Box\bar{h}_{\mu\nu} - \eta_{\mu\nu}\partial^\alpha\partial^\rho\bar{h}_{\alpha\rho} - \frac{1}{2}\eta_{\alpha\nu}\partial^\alpha\partial_\mu\bar{h} \\
&\quad - \frac{1}{2}\eta_{\mu\alpha}\partial^\alpha\partial_\nu\bar{h} + \frac{1}{2}\eta_{\mu\nu}\Box\bar{h} + \partial_\nu\partial_\mu\bar{h} + \frac{1}{2}\eta_{\mu\nu}\eta_{\alpha\rho}\partial^\alpha\partial^\rho\bar{h} - \eta_{\mu\nu}\Box\bar{h}\Big\} \\
&= \frac{1}{2}\left(\partial^\alpha\partial_\mu\bar{h}_{\alpha\nu} + \partial^\alpha\partial_\nu\bar{h}_{\mu\alpha} - \Box\bar{h}_{\mu\nu} - \eta_{\mu\nu}\partial^\alpha\partial^\rho\bar{h}_{\alpha\rho}\right)
\end{aligned} \tag{A.25}$$

Upon imposing the Lorentz gauge $\partial^\nu\bar{h}_{\mu\nu} = 0$ we have

$$G_{\mu\nu} = \frac{1}{2}\Box\bar{h}_{\mu\nu} \tag{A.26}$$

and finally plugging into the EFE and multiplying by 2

$$\Box\bar{h}_{\mu\nu} = \frac{16\pi G}{c^4}T_{\mu\nu} \tag{A.27}$$

which is the wave equation for the gravitational field.

Appendix B

B.1 Phase of a GW signal

We assume that in the rest frame of the neutron star the time dependence of the phase of the gravitational-wave signal can be written as a power series of the form:

$$\Psi_{\rm ns}(\tau) = \Phi_0 + 2\pi \sum_{k=0}^{s} f_{\rm ns}^{(k)} \frac{\tau^{k+1}}{(k+1)!}, \tag{B.1}$$

where τ is the proper time in the neutron star rest frame. The assumption (B.1) means that the instantaneous frequency of the signal in the rest frame of the neutron star is given as

$$f_{\rm ns}(\tau) := \frac{1}{2\pi}\frac{d\Psi_{\rm ns}(\tau)}{d\tau} = \sum_{k=0}^{s} f_{\rm ns}^{(k)} \frac{\tau^k}{k!}, \tag{B.2}$$

so $f_{\rm ns}^{(k)}$ is the kth time derivative of the frequency evaluated at $\tau = 0$. We assume that the neutron star is moving with respect to the SSB uniformly along a straight line according to the equation

$$\boldsymbol{r}_{\rm ns}(t) = r_0 \mathbf{n}_0 + v_{\rm ns}\mathbf{n}_v \left(t + \frac{r_0}{c}\right), \tag{B.3}$$

where $r_0 := |\boldsymbol{r}_{\rm ns}(t = -r_0/c)|, \mathbf{n}_0 := \boldsymbol{r}_{\rm ns}(t = -r_0/c)/r_0$. If we denote by $\boldsymbol{v}_{\rm ns}$ the constant velocity vector of the neutron star then $\boldsymbol{v}_{\rm ns} := |\boldsymbol{v}_{\rm ns}|$ and $\mathbf{n}_v := \boldsymbol{v}_{\rm ns}/v_{\rm ns}$. The time t in Eq. (B.1) is the time coordinate in the SSB rest frame. We do not allow the neutron star to have an intrinsic acceleration. This means we exclude binary neutron stars, except of the binary periods so long that the acceleration effects may be accurately approximated by a Taylor series during the observation time. The phase observed at the SSB at some time t was emitted by the star at the coordinate time t' such that

$$t = t' + \frac{|r_{\rm ns}(t')|}{c}. \tag{B.4}$$

One can show that the relation between the time t' and the star's proper time τ is as follows

$$\tau = \sqrt{1 - \beta_{\rm ns}^2}\left(t' + \frac{r_0}{c}\right), \tag{B.5}$$

where $\beta_{\rm ns} := v_{\rm ns}/c$. In Eq. (B.5) the time dilation effect is taken into account. We have also assumed that $\tau = 0$ when the star's position vector w.r.t. the SSB is $r_0\mathbf{n}_0$. We can now write

$$\Psi_{\rm SSB}(t) = \Psi_{\rm ns}(\tau). \tag{B.6}$$

where $\Psi_{\rm SSB}(t)$ is the phase observed at the SSB at time t, and the time τ can be expressed in terms of t by means of Eqs. (B.4) and (B.6). Collecting Eqs. (B.1) and (B.4)-(B.6) together we can write

$$\Psi_{\rm SSB}(t) = \Phi_0 + 2\pi \sum_{k=0}^{s} f_{\rm ns}^{(k)} \frac{(1-\beta_{\rm ns}^2)^{(k+1)/2}}{(k+1)!} \left(t'(t) + \frac{r_0}{c}\right)^{k+1}, \tag{B.7}$$

where t' is the solution of Eq. (B.4) for a given time t. It reads

$$\begin{aligned} t' = \frac{1}{1-\beta_{\rm ns}^2}\Bigg[& t + \frac{r_0}{c}\beta_{\rm ns}(\beta_{\rm ns} + (\mathbf{n}_0\cdot\mathbf{n}_v)) \\ & - \sqrt{\beta_{\rm ns}^2 t^2 + 2\frac{r_0}{c}\beta_{\rm ns}[\beta_{\rm ns} + (\mathbf{n}_0\cdot\mathbf{n}_u)]t + \frac{r_0^2}{c^2}[1+\beta_{\rm ns}(\mathbf{n}_0\cdot\mathbf{n}_v)]}\Bigg]. \end{aligned} \tag{B.8}$$

We expand the function $\Psi_{\rm SSB}$ given by Eqs. (B.7) and (B.8) w.r.t. time t around $t = 0$. The first few terms of the expansion read

$$\begin{aligned} \frac{\Psi_{\rm SSB}(t) - \Psi_0}{2\pi} = {} & f_{\rm SSB}^0 t + \left\{ f_{\rm SSB}^{(1)} + \frac{((\mathbf{n}_0\cdot\mathbf{n}_v)^2 - 1)\beta_{\rm ns}^2}{(1+(\mathbf{n}_0\cdot\mathbf{n}_v)\beta_{\rm ns})^2 (r_0/c)} f_{\rm SSB}^{(0)} \right\} \frac{t^2}{2} \\ & + \left\{ f_{\rm SSB}^{(2)} + \frac{3((\mathbf{n}_0\cdot\mathbf{n}_v)^2 - 1)\beta_{\rm ns}^2}{(1+(\mathbf{n}_0\cdot\mathbf{n}_v)\beta_{\rm ns})^2 (r_0/c)} \left[f_{\rm SSB}^{(1)} \right.\right. \\ & \left.\left. - \frac{(\beta_{\rm ns} + (\mathbf{n}_0\cdot\mathbf{n}_v))\beta_{\rm ns}^2}{(1+(\mathbf{n}_0\cdot\mathbf{n}_v)\beta_{\rm ns})^2 (r_0/c)} f_{\rm SSB}^{(0)} \right] \right\} \frac{t^3}{6} + \mathcal{O}(t^4), \end{aligned} \tag{B.9}$$

where

$$f_{\rm SSB}^{(k)} := \frac{(1-\beta_{\rm ns}^2)^{\frac{(k+1)}{2}}}{(1+(\mathbf{n}_0\cdot\mathbf{n}_v)\beta_{\rm ns})^{k+1}} f_{\rm ns}^{(k)}, \quad k = 0, \ldots, s. \tag{B.10}$$

As a result of the motion of the neutron star w.r.t. the SSB the Taylor expansion (B.9) of the phase $\Psi_{\rm SSB}$ contains infinitely many terms, even if we restrict, as in Eq. (B.1), the intrinsic spindown of the star to a finite number of terms. When the neutron star moves radially w.r.t. the SSB then $(\mathbf{n}_0\cdot\mathbf{n}_v)^2 = 1$ and the function $\Psi_{\rm SSB}$ can exactly be written as the finite sum:

$$\Psi_{\rm SSB}(t) = \Phi_0 + 2\pi \sum_{k=0}^{s} f_{\rm SSB}^{(k)} \frac{t^{k+1}}{(k+1)!}. \tag{B.11}$$

We shall assume the following polynomial model of the phase of the gravitational radiation observed at the SSB:

$$\Psi_{\rm SSB}(t) = \Phi_0 + 2\pi \sum_{k=0}^{s} f_0^{(k)} \frac{t^{k+1}}{(k+1)!}, \tag{B.12}$$

where the new spindown parameters $f_0^{(k)}$ do not in general coincide with the Doppler scaled intrinsic spindown parameters $f_{\rm SSB}^{(k)}$ defined by Eq. (B.10). We write the position vector r_d of the detector with respect to the SSB as

$$\boldsymbol{r}_d(t) = r_d(t)\mathbf{n}_d(t), \tag{B.13}$$

where $r_d(t) := |r_d(t)|$ and $\mathbf{n}_d(t) := \boldsymbol{r}_d(t)/r_d(t)$. The phase of the gravitational-wave signal at the time t at the detector's location corresponds to the phase near the neutron star at an earlier instant of time t'', where t'' is the solution of the equation

$$t = t'' + \frac{|\boldsymbol{r}_{\rm ns}(t'') - \boldsymbol{r}_d(t)|}{c}. \tag{B.14}$$

The same value of the phase is observed at the SSB at time

$$t'' + \frac{|\boldsymbol{r}_{\rm ns}(t'')|}{c}, \tag{B.15}$$

thus using Eq. (B.10) we can write

$$\Psi_d(t) = \Phi_0 + 2\pi \sum_{k=0}^{s} \frac{f_0^{(k)}}{(K+1)!} \left(t^{''}(t) + \frac{|\boldsymbol{r}_{\rm ns}(t''(t))|}{c} \right)^{k+1} \tag{B.16}$$

where $t''(t)$ is the solution of Eq. (B.14) for given time t. Using Eqs. (B.3) and (B.13) we express the solution t'' to Eq. (124) in terms of the time t and the two small parameters $\beta_{\rm ns}$ and $x := r_d/r_0$:

$$\begin{aligned} t''(x, \beta_{\rm ns}) = \frac{1}{1-\beta_{\rm ns}^2} \Bigg[& t + \frac{r_0}{c}\beta_{\rm ns}[(\mathbf{n}_0 \cdot \mathbf{n}_v) - (\mathbf{n}_d \cdot \mathbf{n}_v)x + \beta_{\rm ns}] \\ & - \Bigg(\left(t + \frac{r_0}{c}\beta_{\rm ns}[(\mathbf{n}_0 \cdot \mathbf{n}_v) - (\mathbf{n}_d \cdot \mathbf{n}_v)x + \beta_{\rm ns}] \right)^2 + (1-\beta_{\rm ns}^2) \\ & \times \left(\frac{r_0^2}{c^2}[1 + \beta_{\rm ns}^2 - 2(\mathbf{n}_0 \cdot \mathbf{n}_d)x + x^2 + 2((\mathbf{n}_0 \cdot \mathbf{n}_v) - (\mathbf{n}_d \cdot \mathbf{n}_v)x)\beta_{\rm ns}] - t^2 \right) \Bigg)^{1/2} \Bigg]. \end{aligned} \tag{B.17}$$

Using Eq. (B.13) we also find that

$$\frac{|\boldsymbol{r}_{\rm ns}(t'')|}{c} = \sqrt{\frac{r_0^2}{c^2} + 2\frac{r_0}{c}\left(t'' + \frac{r_0}{c}\right)(\mathbf{n}_0 \cdot \mathbf{n}_v)\beta_{\rm ns} + \left(t'' + \frac{r_0}{c}\right)^2 \beta_{\rm ns}^2}. \tag{B.18}$$

We now study how to simplify the phase Ψ_d given by Eqs. (B.16)-(B.18).

An optimal method to detect our signal in noise developed in Section 3 involves correlating the data with templates of the signal. In general if the phase of the template differs from that of the signal by as little as 1/4 of a cycle the correlation will be significantly reduced. Thus we adopt the criterion that we exclude an effect from the model of the signal in the case when it contributes less than 1/4 of a cycle to the phase of the signal during the observation time. That this criterion is only a sufficient condition but not necessary follows from the correlations among parameters of the phase. The shifts in the values of the parameters in the template phase away from the true values of the parameters in the signal phase can compensate for the effects in the signal not taken into account in the templates. This effect was observed for the case of coalescing binaries [279, 279, 280]. Finally we stress that such shifts in the template parameter values mean that the estimators of the parameters of the signal when using an inaccurate template will be biased. It may happen that these biases are much larger than the rms errors of the estimators. Thus templates accurate to 1/4 of a cycle over the observation time may not be needed to detect the signal, but they will be needed to obtain accurate estimates of the errors in

parameter measurements. In calculating the number of cycles we assume a long observation time of 120 days, the maximum gravitational wave frequency of 1 kHz, and the extreme case of a neutron star at a distance $r_0 = 40$ pc with $v_{\rm ns} = 10^3$ km/s. For this extreme case the parameters x and $\beta_{\rm ns}$ assume the values (as to a good approximation $r_d \cong 1$ AU):

$$x = 1.21 \times 10^{-7}, \quad \beta_{\rm ns} = 3.34 \times 10^{-3}. \tag{B.19}$$

The numerical values of the spindown parameters $f_0^{(k)}$ we estimate by means of the relation:

$$\left| f_0^{(k)} \right| \cong k! \frac{f_0}{\tau^k}, \tag{B.20}$$

where f_0 is the radiation frequency and τ is the spindown age of the neutron star. As the extreme case we will consider $\tau = 40$ years. It is convenient to carry out the Taylor expansion of the phase (B.16) with respect to the parameters x and $\beta_{\rm ns}$. We note that for any n

$$\left. \frac{\partial^n \Psi_d}{\partial \beta_{\rm ns}^n} \right| (x = 0, \beta_{\rm ns} = 0) = 0. \tag{B.21}$$

Analysis of the first few terms of the Taylor expansion shows that for the observation times $T_o \leq 120$ days, neutron star distances $r_0 \geq 40$ pc, velocities $v_{\rm ns} \leq 10^3$ km/s, frequencies $f_0 \leq 1$ kHz, and spindown ages $\tau \geq 40$ years, the only terms which can contribute more than 1/4 of a cycle to the phase of the signal, read

$$\Psi_d \cong \Phi_0 + 2\pi \sum_{k=0}^{4} f_0^{(k)} \frac{t^{k+1}}{(k+1)!} + \frac{2\pi}{c} \left(\mathbf{n}_0 + \frac{\boldsymbol{v}_{\rm ns\perp}}{r_0} t \right) \cdot \boldsymbol{r}_d \sum_{k=0}^{3} f_0^{(k)} \frac{t^k}{k!}, \tag{B.22}$$

where $\boldsymbol{v}_{\rm ns\perp} := \boldsymbol{v}_{\rm ns} - (\mathbf{n}_0 \cdot \boldsymbol{v}_{\rm ns})\mathbf{n}_0$ is the component of the velocity $\boldsymbol{v}_{\rm ns}$ perpendicular to the vector $\mathbf{n}_0$. The ratio $\boldsymbol{v}_{\rm ns\perp}/r_0$ determines the proper motion of the star. The term in the above expansion proportional to $\boldsymbol{v}_{\rm ns\perp}/r_0$ contributes at most ~ 4 cycles. We shall not consider it in this paper. We shall look at the possibility of its determination in the next paper of this series. Consequently we restrict ourselves to a phase model at the detector of the form

$$\Psi_d \cong \Phi_0 + 2\pi \sum_{k=0}^{4} f_0^{(k)} \frac{t^{k+1}}{(k+1)!} + \frac{2\pi}{c} \mathbf{n}_0 \cdot \boldsymbol{r}_d \sum_{k=0}^{3} f_0^{(k)} \frac{t^k}{k!}. \tag{B.23}$$

The model (B.23) contains the position r_d of the Earth relative to the SSB, which we now consider. In addition we must consider extra, purely relativistic effects left out of (B.23).

Motion of the Earth w.r.t. the SSB is very well determined and there are several computer ephemeris routines available. Here we assume for simplicity that the Earth moves on a circular orbit around the Sun. The eccentricity of the Earth's orbit ($e_\oplus = 0.017$) introduces a change of about 8.3×10^3 cycles in the phase w.r.t. the phase for circular orbit for 1 kHz signal, so it must be included in realistic filters. But it introduces no new parameters so we ignore it here. We also ignore the motion of the Earth around the Earth-Moon barycenter. There are two types of relativistic corrections. One originates in the difference between the coordinate time t which we used in the derivation of the phase model and the proper time

τ in the detector's reference frame. The difference is due to the combined effect of the gravitational redshift and the time dilation. The other correction is the Shapiro delay caused by the propagation of the gravitational wave through the curved spacetime of the solar system. We estimate the contribution to the number of cycles in the phase produced by these corrections.

The difference between the coordinate time t in the first order post-Newtonian coordinate system which is assumed to be the rest frame of the SSB and the proper time τ kept by a terrestrial clock is discussed in detail in Ref. [281]. The difference $\Delta t := t - \tau$ is given by the integral

$$\Delta_E = \frac{1}{c^2}\int_0^t \left[U\left(\boldsymbol{r}(t')\right) + \frac{v(t')^2}{2}\right] dt', \tag{B.24}$$

where r is the position vector of the clock w.r.t. the SSB, $\boldsymbol{v} := \dot{\boldsymbol{r}}$ is the clock's coordinate velocity, and $U\left(\boldsymbol{r}(t')\right)$ is the instantaneous gravitational potential at the clock's location. The time difference described by the integral (B.24) can be split into the secular and periodic part. The secular difference is due to the practically constant rotational velocity and the Earth's gravitational potential at the detector's location as well as the average orbital velocity of the Earth and the average gravitational potential along the Earth's orbit. This secular difference corresponds to the rescaling of the time coordinate and can be incorporated into the definition of the spindown parameters. The main contribution to the periodic part of the integral (B.24) was calculated by Clemence and Szebehely [282] and then corrected by Blandford and Teukolsky [283]. It can be written as

$$(\Delta_E)_{\text{periodic}} \cong \frac{2GM_\odot e_\oplus}{c^2 a_\oplus (1-e_\oplus^2)\Omega_0}\left[\left(1-\frac{1}{8}e_\oplus^2\right)\sin M_\oplus + \frac{1}{2}e_\oplus \sin 2M_\oplus + \frac{3}{8}e_\oplus^2 \sin 3M_\oplus\right], \tag{B.25}$$

where $M_\odot$ is the mass of the Sun, $a_\oplus = 1$ AU, $\Omega_\oplus$ is the mean orbital angular velocity of the Earth, $e_\oplus$ and $M_\oplus$ are the eccentricity and mean anomaly of the Earth's orbit. The quantity $(\Delta t)_{\text{periodic}}$ varies in time with the period of one year and has the amplitude $\sim 1.7 \times 10^{-3}$ s, so for a 1 kHz gravitational wave the contribution of this correction to the total number of cycles is not greater than ~ 2 cycles. Even when it must be included in a filter, it introduces no new parameters. The magnitude of the Shapiro delay can be estimated from the relation [284] (neglecting the eccentricity of the Earth's orbit)

$$\Delta_S = \frac{2GM_\odot}{c^3}\log\left(\frac{1}{1+\cos\theta}\right), \tag{B.26}$$

where θ is the star-Sun-detector angle at the time of observation. To estimate the maximum value of the Shapiro delay we consider a neutron star in such position that at some instant of time the line of sight from the detector to the neutron star is tangent to the surface of the Sun. Then $\theta = \theta_1 \approx \pi - \zeta$, where $\zeta \approx R_\odot/1 \text{ AU} \approx 4.65 \times 10^{-3}$ rad ($R_\odot$ is the radius of the Sun). Six months later $\theta = \theta_2 \approx \zeta$, so the amplitude of the correction is

$$\Delta_S(\theta = \theta_1) - \Delta_S(\theta = \theta_2) \approx \frac{2GM_\odot}{c^3}\log\frac{1+\cos\theta_2}{1+\cos\theta_1} \approx 1.2 \times 10^{-4} \text{ s}. \tag{B.27}$$

For a 1 kHz gravitational wave this gives ~ 0.1 cycles. So the Shapiro delay will be unobservable. We see that the relativistic corrections that need to be applied to our formula are small. By our 1/4 of a

cycle criterion they can be neglected if we search for signals with frequencies less than ~ 100 Hz. We shall not consider these corrections in this and the following papers of the series since they are unlikely to influence our results. However they should be included in filters.

Bibliography

[1] J. Weber. Detection and generation of gravitational waves. *Phys. Rev.*, 117:306–313, Jan 1960. doi: 10.1103/PhysRev.117.306. URL https://link.aps.org/doi/10.1103/PhysRev.117.306.

[2] J. H. Taylor, L. A. Fowler, and P. M. McCulloch. Measurements of general relativistic effects in the binary pulsar psr1913 + 16. *Nature*, 277(5696):437–440, Feb 1979. ISSN 1476-4687. doi: 10.1038/277437a0. URL https://doi.org/10.1038/277437a0.

[3] B. P. Abbott et al. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 116:061102, Feb 2016. doi: 10.1103/PhysRevLett.116.061102. URL https://link.aps.org/doi/10.1103/PhysRevLett.116.061102.

[4] B. P. Abbott et al. Multi-messenger observations of a binary neutron star merger*. *The Astrophysical Journal Letters*, 848(2):L12, oct 2017. doi: 10.3847/2041-8213/aa91c9. URL https://dx.doi.org/10.3847/2041-8213/aa91c9.

[5] B. P. Abbott et al. Multi-messenger observations of a binary neutron star merger. *ApJL*, 848(2): L12, oct 2017. doi: 10.3847/2041-8213/aa91c9. URL https://dx.doi.org/10.3847/2041-8213/aa91c9.

[6] J. B. Griffiths and J. Podolský. *Exact Space-Times in Einstein's General Relativity*. Cambridge Monographs on Mathematical Physics. Cambridge University Press, 2009.

[7] C.W. Misner, K.S. Thorne, J.A. Wheeler, and D.I. Kaiser. *Gravitation*. Princeton University Press, 2017. ISBN 9781400889099. URL https://books.google.co.jp/books?id=zAAuDwAAQBAJ.

[8] P.M. McHugh et al. Calibration of the allegro resonant detector. *Classical and Quantum Gravity*, 22(18):S965, aug 2005. doi: 10.1088/0264-9381/22/18/S10. URL https://dx.doi.org/10.1088/0264-9381/22/18/S10.

[9] Vinante A. and the AURIGA Collaboration. Present performance and future upgrades of the auriga capacitive readout. *Classical and Quantum Gravity*, 23(8):S103, mar 2006. doi: 10.1088/0264-9381/23/8/S14. URL https://dx.doi.org/10.1088/0264-9381/23/8/S14.

[10] P Astone et al. Status report on the explorer and nautilus detectors and the present science run. *Classical and Quantum Gravity*, 23(8):S57, mar 2006. doi: 10.1088/0264-9381/23/8/S08. URL https://dx.doi.org/10.1088/0264-9381/23/8/S08.

[11] D. G. Blair et al. High sensitivity gravitational wave antenna with parametric transducer readout. *Phys. Rev. Lett.*, 74:1908–1911, Mar 1995. doi: 10.1103/PhysRevLett.74.1908. URL https://link.aps.org/doi/10.1103/PhysRevLett.74.1908.

[12] L. Gottardi et al. Sensitivity of the spherical gravitational wave detector minigrail operating at 5 k. *Phys. Rev. D*, 76:102005, Nov 2007. doi: 10.1103/PhysRevD.76.102005. URL https://link.aps.org/doi/10.1103/PhysRevD.76.102005.

[13] O D Aguiar et al. The brazilian gravitational wave detector mario schenberg: status report. *Classical and Quantum Gravity*, 23(8):S239, mar 2006. doi: 10.1088/0264-9381/23/8/S30. URL https://dx.doi.org/10.1088/0264-9381/23/8/S30.

[14] Aaron D. Johnson et al. NANOGrav 15-year gravitational-wave background methods. *Phys. Rev. D*, 109(10):103012, 2024. doi: 10.1103/PhysRevD.109.103012.

[15] M. Colpi et al. Lisa definition study report, 2024. URL https://arxiv.org/abs/2402.07571.

[16] S. Kawamura et al. The japanese space gravitational wave antenna—decigo. *Classical and Quantum Gravity*, 23(8):S125, mar 2006. doi: 10.1088/0264-9381/23/8/S17. URL https://dx.doi.org/10.1088/0264-9381/23/8/S17.

[17] F. Jenet et al. The north american nanohertz observatory for gravitational waves, 2009. URL https://arxiv.org/abs/0909.1058.

[18] G. Hobbs et al. The pulse@parkes project: a new observing technique for long-term pulsar monitoring. *Publications of the Astronomical Society of Australia*, 26(4):468–475, 2009. doi: 10.1071/AS09021. URL https://doi.org/10.1071/AS09021.

[19] G Hobbs et al. The international pulsar timing array project: using pulsars as a gravitational wave detector. *Classical and Quantum Gravity*, 27(8):084013, apr 2010. doi: 10.1088/0264-9381/27/8/084013. URL https://dx.doi.org/10.1088/0264-9381/27/8/084013.

[20] R D Ferdman et al. The european pulsar timing array: current efforts and a leap toward the future. *Classical and Quantum Gravity*, 27(8):084014, apr 2010. doi: 10.1088/0264-9381/27/8/084014. URL https://dx.doi.org/10.1088/0264-9381/27/8/084014.

[21] K. Danzmann. Lisa — an esa cornerstone mission for the detection and observation of gravitational waves. *Advances in Space Research*, 32(7):1233–1242, 2003. ISSN 0273-1177. doi: https://doi.org/10.1016/S0273-1177(03)90323-1. URL https://www.sciencedirect.com/science/article/pii/S0273117703903231. Fundamental Physics in Space.

[22] The LIGO Scientific Collaboration, J Aasi, et al. Advanced ligo. *Classical and Quantum Gravity*, 32(7):074001, mar 2015. doi: 10.1088/0264-9381/32/7/074001. URL https://dx.doi.org/10.1088/0264-9381/32/7/074001.

[23] F. Acernese et al. Advanced virgo: a second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, 32(2):024001, dec 2014. doi: 10.1088/0264-9381/32/2/024001. URL https://dx.doi.org/10.1088/0264-9381/32/2/024001.

[24] H. Abe et al. The current status and future prospects of kagra, the large-scale cryogenic gravitational wave telescope built in the kamioka underground. *Galaxies*, 10(3), 2022. ISSN 2075-4434. doi: 10.3390/galaxies10030063. URL https://www.mdpi.com/2075-4434/10/3/63.

[25] M. Ando et al. Stable operation of a 300-m laser interferometer with sufficient sensitivity to detect gravitational-wave events within our galaxy. *Phys. Rev. Lett.*, 86:3950–3954, Apr 2001. doi: 10.1103/PhysRevLett.86.3950. URL https://link.aps.org/doi/10.1103/PhysRevLett.86.3950.

[26] K. Kuroda and the LCGT Collaboration. The status of lcgt. *Classical and Quantum Gravity*, 23(8):S215, mar 2006. doi: 10.1088/0264-9381/23/8/S27. URL https://dx.doi.org/10.1088/0264-9381/23/8/S27.

[27] K L Dooley and (for the LIGO Scientific Collaboration). Status of geo 600. *Journal of Physics: Conference Series*, 610(1):012015, apr 2015. doi: 10.1088/1742-6596/610/1/012015. URL https://dx.doi.org/10.1088/1742-6596/610/1/012015.

[28] P. Barriga et al. Aigo: A southern hemisphere detector for the worldwide array of ground-based interferometric gravitational wave detectors. *Classical and Quantum Gravity*, 07 2013. doi: 10.1088/0264-9381/27/8/084005.

[29] Chiummo, A. The einstein telescope: Status of the project. *EPJ Web Conf.*, 280:03003, 2023. doi: 10.1051/epjconf/202328003003. URL https://doi.org/10.1051/epjconf/202328003003.

[30] Evan D. Hall. Cosmic explorer: A next-generation ground-based gravitational-wave observatory. *Galaxies*, 10(4), 2022. ISSN 2075-4434. doi: 10.3390/galaxies10040090. URL https://www.mdpi.com/2075-4434/10/4/90.

[31] P. Astone et al. Methods and results of the igec search for burst gravitational waves in the years 1997–2000. *Phys. Rev. D*, 68:022001, Jul 2003. doi: 10.1103/PhysRevD.68.022001. URL https://link.aps.org/doi/10.1103/PhysRevD.68.022001.

[32] P. Astone et al. Igec2: A 17-month search for gravitational wave bursts in 2005–2007. *Phys. Rev. D*, 82:022003, Jul 2010. doi: 10.1103/PhysRevD.82.022003. URL https://link.aps.org/doi/10.1103/PhysRevD.82.022003.

[33] S. Husa. Michele maggiore: Gravitational waves. volume 1: theory and experiments. *General Relativity and Gravitation*, 41(7):1667–1669, Jul 2009. ISSN 1572-9532. doi: 10.1007/s10714-009-0762-5. URL https://doi.org/10.1007/s10714-009-0762-5.

[34] A.E. Siegman. *Lasers*. G - Reference,Information and Interdisciplinary Subjects Series. University Science Books, 1986. ISBN 9780935702118. URL https://books.google.co.jp/books?id=1BZVwUZLTkAC.

[35] S. Hild. *A Basic Introduction to Quantum Noise and Quantum-Non-Demolition Techniques*, pages 291–314. Springer International Publishing, Cham, 2014. ISBN 978-3-319-03792-9. doi: 10.1007/978-3-319-03792-9_11. URL https://doi.org/10.1007/978-3-319-03792-9_11.

[36] KAGRA Collaboration. Latest estimated sensitivity of kagra (v201708). Tech. Rep. T1707038-v9, JGW Document, 2017. URL https://gwdoc.icrr.u-tokyo.ac.jp/cgi-bin/DocDB/ShowDocument?docid=7038.

[37] LIGO Scientific Collaboration. The lsc-virgo-kagra observational science white paper. Technical report, LIGO Document Control Center, 2020. URL https://dcc.ligo.org/LIGO-T2000424/public.

[38] B. S. Sathyaprakash and B. F. Schutz. Physics, astrophysics and cosmology with gravitational waves. *Living Reviews in Relativity*, 12(1):2, Dec 2009. ISSN 1433-8351. doi: 10.12942/lrr-2009-2. URL https://doi.org/10.12942/lrr-2009-2.

[39] T. Damour, R. Jantzen, and R. Ruffini. *The Twelfth Marcel Grossmann Meeting*. WORLD SCIENTIFIC, 2012. doi: 10.1142/8316. URL https://www.worldscientific.com/doi/abs/10.1142/8316.

[40] J Abadie, others, (The LIGO Scientific Collaboration, the Virgo Collaboration), and K Belczynski. Predictions for the rates of compact binary coalescences observable by ground-based gravitational-wave detectors. *Classical and Quantum Gravity*, 27(17):173001, jul 2010. doi: 10.1088/0264-9381/27/17/173001. URL https://dx.doi.org/10.1088/0264-9381/27/17/173001.

[41] M. Campanelli, C. O. Lousto, P. Marronetti, and Y. Zlochower. Accurate evolutions of orbiting black-hole binaries without excision. *Phys. Rev. Lett.*, 96:111101, Mar 2006. doi: 10.1103/PhysRevLett.96.111101. URL https://link.aps.org/doi/10.1103/PhysRevLett.96.111101.

[42] B. F. Schutz. *Sources of Gravitational Radiation*, pages 3–17. Springer Netherlands, Dordrecht, 1989. ISBN 978-94-009-1185-7. doi: 10.1007/978-94-009-1185-7_1. URL https://doi.org/10.1007/978-94-009-1185-7_1.

[43] S. E. Gossan et al. Observing gravitational waves from core-collapse supernovae in the advanced detector era. *Phys. Rev. D*, 93:042002, Feb 2016. doi: 10.1103/PhysRevD.93.042002. URL https://link.aps.org/doi/10.1103/PhysRevD.93.042002.

[44] V. Radhakrishnan. *Pulsar Velocities and Their Origins*, pages 445–452. Springer Netherlands, Dordrecht, 1992. doi: 10.1007/978-94-011-2704-2_39. URL https://doi.org/10.1007/978-94-011-2704-2_39.

[45] C. D. Ott et al. A new mechanism for gravitational-wave emission in core-collapse supernovae. *Phys. Rev. Lett.*, 96:201102, May 2006. doi: 10.1103/PhysRevLett.96.201102. URL https://link.aps.org/doi/10.1103/PhysRevLett.96.201102.

[46] S. W. Hawking and W. Israel. *Three Hundred Years of Gravitation*. 1989.

[47] V.M. Kaspi. Grand unification of neutron stars. *Proceedings of the National Academy of Sciences*, 107(16):7147–7152, 2010. doi: 10.1073/pnas.1000812107. URL https://www.pnas.org/doi/abs/10.1073/pnas.1000812107.

[48] K. Hurley et al. An exceptionally bright flare from sgr 1806–20 and the origins of short-duration γ-ray bursts. *Nature*, 434(7037):1098–1103, Apr 2005. ISSN 1476-4687. doi: 10.1038/nature03519. URL https://doi.org/10.1038/nature03519.

[49] U. S. Inan et al. Massive disturbance of the daytime lower ionosphere by the giant γ-ray flare from magnetar sgr 1806–20. *Geophysical Research Letters*, 34(8), 2007. doi: https://doi.org/10.1029/2006GL029145. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GL029145.

[50] L. Leblond et al. Gravitational waves from broken cosmic strings: The bursts and the beads. *Phys. Rev. D*, 79:123519, Jun 2009. doi: 10.1103/PhysRevD.79.123519. URL https://link.aps.org/doi/10.1103/PhysRevD.79.123519.

[51] P. Brady et al. Searching for periodic sources with ligo. *Phys. Rev. D*, 57:2101–2116, Feb 1998. doi: 10.1103/PhysRevD.57.2101. URL https://link.aps.org/doi/10.1103/PhysRevD.57.2101.

[52] C. Fryer and S. Woosley. A new twist on neutron stars. *Nature*, 411(6833):31–33, May 2001. ISSN 1476-4687. doi: 10.1038/35075184. URL https://doi.org/10.1038/35075184.

[53] L. P. Grishchuk. Amplification of gravitational waves in an istropic universe. *Zh. Eksp. Teor. Fiz.*, 67:825–838, 1974.

[54] Arianna I. Renzini, Boris Goncharov, Alexander C. Jenkins, and Patrick M. Meyers. Stochastic gravitational-wave backgrounds: Current detection efforts and future prospects. *Galaxies*, 10(1), 2022. ISSN 2075-4434. doi: 10.3390/galaxies10010034. URL https://www.mdpi.com/2075-4434/10/1/34.

[55] S. Bose et al. Towards the first search for a stochastic background in LIGO data: Applications of signal simulations. *Class. Quant. Grav.*, 20:S677–S687, 2003. doi: 10.1088/0264-9381/20/17/311.

[56] J. Beringer et al. Review of particle physics. *Phys. Rev. D*, 86:010001, Jul 2012. doi: 10.1103/PhysRevD.86.010001. URL https://link.aps.org/doi/10.1103/PhysRevD.86.010001.

[57] B. Allen and J. D. Romano. Detecting a stochastic background of gravitational radiation: Signal processing strategies and sensitivities. *Phys. Rev. D*, 59:102001, Mar 1999. doi: 10.1103/PhysRevD.59.102001. URL https://link.aps.org/doi/10.1103/PhysRevD.59.102001.

[58] A. A. Starobinsky. Spectrum of relict gravitational radiation and the early state of the universe. *JETP Lett.*, 30:682–685, 1979.

[59] Rennan Bar-Kana. Limits on direct detection of gravitational waves. *Phys. Rev. D*, 50:1157–1160, Jul 1994. doi: 10.1103/PhysRevD.50.1157. URL https://link.aps.org/doi/10.1103/PhysRevD.50.1157.

[60] Edward W. Kolb. *The Early Universe*, volume 69. Taylor and Francis, 5 2019. ISBN 978-0-429-49286-0, 978-0-201-62674-2. doi: 10.1201/9780429492860.

[61] T. L. Smith, E. Pierpaoli, and M. Kamionkowski. New cosmic microwave background constraint to primordial gravitational waves. *Phys. Rev. Lett.*, 97:021301, Jul 2006. doi: 10.1103/PhysRevLett.97.021301. URL https://link.aps.org/doi/10.1103/PhysRevLett.97.021301.

[62] Z.H. Zhu et al. Determining the cosmic equation of state using future gravitational wave detectors. *A&A*, 372(2):377–380, 2001. doi: 10.1051/0004-6361:20010458. URL https://doi.org/10.1051/0004-6361:20010458.

[63] C. Wu, V. Mandic, and T. Regimbau. Accessibility of the gravitational-wave background due to binary coalescences to second and third generation gravitational-wave detectors. *Phys. Rev. D*, 85:104024, May 2012. doi: 10.1103/PhysRevD.85.104024. URL https://link.aps.org/doi/10.1103/PhysRevD.85.104024.

[64] V. R. Pandharipande, D. Pines, and R. A. Smith. Neutron star structure: theory, observation, and speculation. *Astrophys. J.*, 208:550–566, September 1976. doi: 10.1086/154637.

[65] M. Zimmermann. Revised estimate of gravitational radiation from crab and vela pulsars. *Nature*, 271(5645):524–525, Feb 1978. ISSN 1476-4687. doi: 10.1038/271524a0. URL https://doi.org/10.1038/271524a0.

[66] B. Haskell, M. Priymak, A. Patruno, M. Oppenoorth, A. Melatos, and P. D. Lasky. Detecting gravitational waves from mountains on neutron stars in the advanced detector era. *Monthly Notices of the Royal Astronomical Society*, 450(3):2393–2403, 05 2015. ISSN 0035-8711. doi: 10.1093/mnras/stv726. URL https://doi.org/10.1093/mnras/stv726.

[67] P. D. Lasky. Gravitational Waves from Neutron Stars: A Review. *Pubs. Astron. Soc. Australia*, 32:e034, September 2015. doi: 10.1017/pasa.2015.35.

[68] K. Glampedakis and L. Gualtieri. *Gravitational Waves from Single Neutron Stars: An Advanced Detector Era Survey*, pages 673–736. Springer International Publishing, Cham, 2018. ISBN 978-3-319-97616-7. doi: 10.1007/978-3-319-97616-7_12. URL https://doi.org/10.1007/978-3-319-97616-7_12.

[69] N. K. Johnson-McDaniel and B. J. Owen. Maximum elastic deformations of relativistic stars. *Phys. Rev. D*, 88:044004, Aug 2013. doi: 10.1103/PhysRevD.88.044004. URL https://link.aps.org/doi/10.1103/PhysRevD.88.044004.

[70] C. J. Horowitz and K. Kadau. Breaking strain of neutron star crust and gravitational waves. *Phys. Rev. Lett.*, 102:191102, May 2009. doi: 10.1103/PhysRevLett.102.191102. URL https://link.aps.org/doi/10.1103/PhysRevLett.102.191102.

[71] N. Andersson. A new class of unstable modes of rotating relativistic stars. *The Astrophysical Journal*, 502(2):708, aug 1998. doi: 10.1086/305919. URL https://dx.doi.org/10.1086/305919.

[72] J. Friedman and S. Morsink. Axial instability of rotating relativistic stars. *The Astrophysical Journal*, 502(2):714, aug 1998. doi: 10.1086/305920. URL https://dx.doi.org/10.1086/305920.

[73] M. Alford and K. Schwenzer. Gravitational wave emission and spin-down of young pulsars. *The Astrophysical Journal*, 781(1):26, jan 2014. doi: 10.1088/0004-637X/781/1/26. URL https://dx.doi.org/10.1088/0004-637X/781/1/26.

[74] M. G. Alford and K. Schwenzer. Gravitational wave emission from oscillating millisecond pulsars. *Monthly Notices of the Royal Astronomical Society*, 446(4):3631–3641, 12 2014. ISSN 0035-8711. doi: 10.1093/mnras/stu2361. URL https://doi.org/10.1093/mnras/stu2361.

[75] S. Chandrasekhar. Solutions of two problems in the theory of gravitational radiation. *Phys. Rev. Lett.*, 24:611–615, Mar 1970. doi: 10.1103/PhysRevLett.24.611. URL https://link.aps.org/doi/10.1103/PhysRevLett.24.611.

[76] L. Lindblom and S. L. Detweiler. The quadrupole oscillations of neutron stars. *Astrophys. J.*, 53: 73–92, September 1983. doi: 10.1086/190884.

[77] D. Doneva, K. Kokkotas, and P. Pnigouras. Gravitational wave afterglow in binary neutron star mergers. *Phys. Rev. D*, 92:104040, Nov 2015. doi: 10.1103/PhysRevD.92.104040. URL https://link.aps.org/doi/10.1103/PhysRevD.92.104040.

[78] B. J. Owen. Maximum elastic deformations of compact stars with exotic equations of state. *Phys. Rev. Lett.*, 95:211101, Nov 2005. doi: 10.1103/PhysRevLett.95.211101. URL https://link.aps.org/doi/10.1103/PhysRevLett.95.211101.

[79] R. Narayan. The Birthrate and Initial Spin Period of Single Radio Pulsars. *Astrophys. J.*, 319:162, August 1987. doi: 10.1086/165442.

[80] M. A. Ruderman and P. G. Sutherland. Theory of pulsars: polar gaps, sparks, and coherent microwave radiation. *Astrophys. J.*, 196:51–72, February 1975. doi: 10.1086/153393.

[81] B. Zhang et al. Radio pulsar death line revisited: Is psr j2144–3933 anomalous? *The Astrophysical Journal*, 531(2):L135, feb 2000. doi: 10.1086/312542. URL https://dx.doi.org/10.1086/312542.

[82] A.G. Lyne and F. Graham-Smith. *Pulsar Astronomy*. Cambridge Astrophysics. Cambridge University Press, 2006. ISBN 9780521839549. URL https://books.google.co.jp/books?id=AK9N3zxL4ToC.

[83] F. Hoyle and R. A. Lyttleton. The effect of interstellar matter on climatic variation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 35(3):405–415, 1939. doi: 10.1017/S0305004100021150.

[84] J. Hessels et al. A Radio Pulsar Spinning at 716 Hz. *Science*, 311(5769):1901–1904, March 2006. doi: 10.1126/science.1123430.

[85] F. C. Michel. Relativistic Stellar-Wind Torques. *Astrophys. J.*, 158:727, November 1969. doi: 10.1086/150233.

[86] B. M. Gaensler and P. O. Slane. The evolution and structure of pulsar wind nebulae. *Annual Review of Astronomy and Astrophysics*, 44(Volume 44, 2006):17–47, 2006. ISSN 1545-4282. doi: https://doi.org/10.1146/annurev.astro.44.051905.092528. URL https://www.annualreviews.org/content/journals/10.1146/annurev.astro.44.051905.092528.

[87] I. Contopoulos, D. Kazanas, and C. Fendt. The axisymmetric pulsar magnetosphere. *The Astrophysical Journal*, 511(1):351, jan 1999. doi: 10.1086/306652. URL https://dx.doi.org/10.1086/306652.

[88] A.K. Harding et al. Magnetar spin-down. *The Astrophysical Journal*, 525(2):L125, nov 1999. doi: 10.1086/312339. URL https://dx.doi.org/10.1086/312339.

[89] F.C. Michel and H. Li. Electrodynamics of neutron stars. *Physics Reports*, 318(6):227–297, 1999. ISSN 0370-1573. doi: https://doi.org/10.1016/S0370-1573(99)00002-2. URL https://www.sciencedirect.com/science/article/pii/S0370157399000022.

[90] A. Melatos. Spin-down of an oblique rotator with a current-starved outer magnetosphere. *Monthly Notices of the Royal Astronomical Society*, 288(4):1049–1059, 07 1997. ISSN 0035-8711. doi: 10.1093/mnras/288.4.1049. URL https://doi.org/10.1093/mnras/288.4.1049.

[91] F. PACINI. Rotating neutron stars, pulsars and supernova remnants. *Nature*, 219(5150):145–146, Jul 1968. ISSN 1476-4687. doi: 10.1038/219145a0. URL https://doi.org/10.1038/219145a0.

[92] A. Andersson. A new class of unstable modes of rotating relativistic stars. *The Astrophysical Journal*, 502(2):708, aug 1998. doi: 10.1086/305919. URL https://dx.doi.org/10.1086/305919.

[93] J. L. Friedman and B. F. Schutz. Secular instability of rotating Newtonian stars. *Astrophys. J.*, 222:281–296, May 1978. doi: 10.1086/156143.

[94] L. Bildsten. Gravitational radiation and rotation of accreting neutron stars. *The Astrophysical Journal*, 501(1):L89, jun 1998. doi: 10.1086/311440. URL https://dx.doi.org/10.1086/311440.

[95] B. J. Owen. How to adapt broad-band gravitational-wave searches for r-modes. *Phys. Rev. D*, 82:104002, Nov 2010. doi: 10.1103/PhysRevD.82.104002. URL https://link.aps.org/doi/10.1103/PhysRevD.82.104002.

[96] K. S. Thorne. Multipole expansions of gravitational radiation. *Rev. Mod. Phys.*, 52:299–339, Apr 1980. doi: 10.1103/RevModPhys.52.299. URL https://link.aps.org/doi/10.1103/RevModPhys.52.299.

[97] C. Palomba. Simulation of a population of isolated neutron stars evolving through the emission of gravitational waves. *Monthly Notices of the Royal Astronomical Society*, 359(3):1150–1164, 05 2005. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2005.08975.x. URL https://doi.org/10.1111/j.1365-2966.2005.08975.x.

[98] B. P. Abbott et al. Prospects for observing and localizing gravitational-wave transients with advanced ligo, advanced virgo and kagra. *Living Reviews in Relativity*, 23(1):3, Sep 2020. ISSN 1433-8351. doi: 10.1007/s41114-020-00026-9. URL https://doi.org/10.1007/s41114-020-00026-9.

[99] K. Riles. Searches for continuous-wave gravitational radiation. *Living Reviews in Relativity*, 26 (1):3, Apr 2023. ISSN 1433-8351. doi: 10.1007/s41114-023-00044-3. URL https://doi.org/10.1007/s41114-023-00044-3.

[100] M. Maggiore et al. Science case for the einstein telescope. *Journal of Cosmology and Astroparticle Physics*, 2020(03):050, mar 2020. doi: 10.1088/1475-7516/2020/03/050. URL https://dx.doi.org/10.1088/1475-7516/2020/03/050.

[101] B.P. Abbott, , others, (LIGO Scientific Collaboration), and J Harms. Exploring the sensitivity of next generation gravitational wave detectors. *Classical and Quantum Gravity*, 34(4):044001, jan 2017. doi: 10.1088/1361-6382/aa51f4. URL https://dx.doi.org/10.1088/1361-6382/aa51f4.

[102] D. Davis et al. Ligo detector characterization in the second and third observing runs. *Classical and Quantum Gravity*, 38(13):135014, jun 2021. doi: 10.1088/1361-6382/abfd85. URL https://dx.doi.org/10.1088/1361-6382/abfd85.

[103] A. Worley, P.G. Krastev, and B. Li. Nuclear constraints on the moments of inertia of neutron stars. *The Astrophysical Journal*, 685(1):390, sep 2008. doi: 10.1086/589823. URL https://dx.doi.org/10.1086/589823.

[104] A. Arvanitaki et al. String axiverse. *Phys. Rev. D*, 81:123530, Jun 2010. doi: 10.1103/PhysRevD.81.123530. URL https://link.aps.org/doi/10.1103/PhysRevD.81.123530.

[105] R. Penrose. Gravitational collapse: The role of general relativity. *Riv. Nuovo Cim.*, 1:252–276, 1969. doi: 10.1023/A:1016578408204.

[106] Ya. B. Zel'Dovich. Generation of Waves by a Rotating Body. *Soviet Journal of Experimental and Theoretical Physics Letters*, 14:180, August 1971.

[107] A. Arvanitaki et al. Discovering the qcd axion with black holes and gravitational waves. *Phys. Rev. D*, 91:084011, Apr 2015. doi: 10.1103/PhysRevD.91.084011. URL https://link.aps.org/doi/10.1103/PhysRevD.91.084011.

[108] J. Silk et al. *Particle Dark Matter: Observations, Models and Searches*. Cambridge Univ. Press, Cambridge, 2010. ISBN 978-1-107-65392-4. doi: 10.1017/CBO9780511770739.

[109] A. Arvanitaki et al. Black hole mergers and the qcd axion at advanced ligo. *Phys. Rev. D*, 95: 043001, Feb 2017. doi: 10.1103/PhysRevD.95.043001. URL https://link.aps.org/doi/10.1103/PhysRevD.95.043001.

[110] R. Brito et al. Stochastic and resolvable gravitational waves from ultralight bosons. *Phys. Rev. Lett.*, 119:131101, Sep 2017. doi: 10.1103/PhysRevLett.119.131101. URL https://link.aps.org/doi/10.1103/PhysRevLett.119.131101.

[111] S. Bonazzola and E. Gourgoulhon. Gravitational waves from pulsars: Emission by the magnetic field induced distortion. *Astron. Astrophys.*, 312:675, 1996.

[112] B. Allen. Gravitational wave detector sites. 7 1996.

[113] M. Zimmermann and E. Szedenits. Gravitational waves from rotating and precessing rigid bodies: Simple models and applications to pulsars. *Phys. Rev. D*, 20:351–355, Jul 1979. doi: 10.1103/PhysRevD.20.351. URL https://link.aps.org/doi/10.1103/PhysRevD.20.351.

[114] B. F. Schutz. The detection of gravitational waves. In D. G. Blair, editor, *The Detection of Gravitational Waves*, pages 406–452. Cambridge University Press, Cambridge, 1991.

[115] P. Jaranowski, A. Królak, and B. F. Schutz. Data analysis of gravitational-wave signals from spinning neutron stars: The signal and its detection. *Phys. Rev. D*, 58:063001, Aug 1998. doi: 10.1103/PhysRevD.58.063001. URL https://link.aps.org/doi/10.1103/PhysRevD.58.063001.

[116] R. Prix. The f-statistic and its implementation in computefstatistic v2. Technical report, LIGO Report T0900149-v6, 2018. URL https://dcc.ligo.org/T0900149.

[117] P. Jaranowski and A. Królak. Searching for gravitational waves from known pulsars using the and statistics. *Classical and Quantum Gravity*, 27(19):194015, sep 2010. doi: 10.1088/0264-9381/27/19/194015. URL https://dx.doi.org/10.1088/0264-9381/27/19/194015.

[118] R. Prix and B. Krishnan. Targeted search for continuous gravitational waves: Bayesian versus maximum-likelihood statistics. *Classical and Quantum Gravity*, 26(20):204013, oct 2009. doi: 10.1088/0264-9381/26/20/204013. URL https://dx.doi.org/10.1088/0264-9381/26/20/204013.

[119] K. Riles. Gravitational waves: Sources, detectors and searches. *Progress in Particle and Nuclear Physics*, 68:1–54, 2013. ISSN 0146-6410. doi: https://doi.org/10.1016/j.ppnp.2012.08.001. URL https://www.sciencedirect.com/science/article/pii/S0146641012001093.

[120] B. Behnke, M. A. Papa, and R. Prix. Postprocessing methods used in the search for continuous gravitational-wave signals from the galactic center. *Phys. Rev. D*, 91:064007, Mar 2015. doi: 10.1103/PhysRevD.91.064007. URL https://link.aps.org/doi/10.1103/PhysRevD.91.064007.

[121] C. Dreissigacker, R. Prix, and K. Wette. Fast and accurate sensitivity estimation for continuous-gravitational-wave searches. *Phys. Rev. D*, 98:084058, Oct 2018. doi: 10.1103/PhysRevD.98.084058. URL https://link.aps.org/doi/10.1103/PhysRevD.98.084058.

[122] B. Abbott et al. All-sky search for periodic gravitational waves in ligo s4 data. *Phys. Rev. D*, 77:022001, Jan 2008. doi: 10.1103/PhysRevD.77.022001. URL https://link.aps.org/doi/10.1103/PhysRevD.77.022001.

[123] B. Steltner et al. Identification and removal of non-gaussian noise transients for gravitational-wave searches. *Phys. Rev. D*, 105:022005, Jan 2022. doi: 10.1103/PhysRevD.105.022005. URL https://link.aps.org/doi/10.1103/PhysRevD.105.022005.

[124] R. Balasubramanian, B. S. Sathyaprakash, and S. V. Dhurandhar. Gravitational waves from coalescing binaries: Detection strategies and monte carlo estimation of parameters. *Phys. Rev. D*, 53:3033–3055, Mar 1996. doi: 10.1103/PhysRevD.53.3033. URL https://link.aps.org/doi/10.1103/PhysRevD.53.3033.

[125] R. Wette, K.and Prix. Flat parameter-space metric for all-sky searches for gravitational-wave pulsars. *Phys. Rev. D*, 88:123005, Dec 2013. doi: 10.1103/PhysRevD.88.123005. URL https://link.aps.org/doi/10.1103/PhysRevD.88.123005.

[126] R. Prix. Template-based searches for gravitational waves: efficient lattice covering of flat parameter spaces. *Classical and Quantum Gravity*, 24(19):S481, sep 2007. doi: 10.1088/0264-9381/24/19/S11. URL https://dx.doi.org/10.1088/0264-9381/24/19/S11.

[127] R. Prix. Search for continuous gravitational waves: Metric of the multidetector $\mathcal{F}$-statistic. *Phys. Rev. D*, 75:023004, Jan 2007. doi: 10.1103/PhysRevD.75.023004. URL https://link.aps.org/doi/10.1103/PhysRevD.75.023004.

[128] K. Wette et al. Searching for gravitational waves from cassiopeia a with ligo. *Classical and Quantum Gravity*, 25(23):235011, nov 2008. doi: 10.1088/0264-9381/25/23/235011. URL https://dx.doi.org/10.1088/0264-9381/25/23/235011.

[129] D. M. Whitbeck. *Observational Consequences of Gravitational Wave Emission From Spinning Compact Sources*. PhD thesis, Pennsylvania State University, 2010. https://etda.libraries.psu.edu/catalog/7132.

[130] R. Prix and M. Shaltev. Search for continuous gravitational waves: Optimal stackslide method at fixed computing cost. *Phys. Rev. D*, 85:084010, Apr 2012. doi: 10.1103/PhysRevD.85.084010. URL https://link.aps.org/doi/10.1103/PhysRevD.85.084010.

[131] P. B. Covas et al. Identification and mitigation of narrow spectral artifacts that degrade searches for persistent gravitational waves in the first two observing runs of advanced ligo. *Phys. Rev. D*, 97:082002, Apr 2018. doi: 10.1103/PhysRevD.97.082002. URL https://link.aps.org/doi/10.1103/PhysRevD.97.082002.

[132] K. Wette. Searches for continuous gravitational waves from neutron stars: A twenty-year retrospective. *Astroparticle Physics*, 153:102880, 2023. ISSN 0927-6505. doi: https://doi.org/10.1016/j.astropartphys.2023.102880. URL https://www.sciencedirect.com/science/article/pii/S092765052300066X.

[133] B. Allen et al. SFT Data Format Version 2–3 Specification. Tech. Rep. T040164-v2, LIGO, 2022. URL https://dcc.ligo.org/LIGO-T040164-v2/public.

[134] P. Astone, S. Frasca, and C. Palomba. The short fft database and the peak map for the hierarchical search of periodic sources. *Classical and Quantum Gravity*, 22(18):S1197, sep 2005. doi: 10.1088/0264-9381/22/18/S34. URL https://dx.doi.org/10.1088/0264-9381/22/18/S34.

[135] P. Astone, S. D'Antonio, S. Frasca, and C. Palomba. A method for detection of known sources of continuous gravitational wave signals in non-stationary data. *Classical and Quantum Gravity*, 27(19):194016, sep 2010. doi: 10.1088/0264-9381/27/19/194016. URL https://dx.doi.org/10.1088/0264-9381/27/19/194016.

[136] P. Astone, A. Colla, S. D'Antonio, S. Frasca, and C. Palomba. Coherent search of continuous gravitational wave signals: extension of the 5-vectors method to a network of detectors. *Journal of Physics: Conference Series*, 363(1):012038, jun 2012. doi: 10.1088/1742-6596/363/1/012038. URL https://dx.doi.org/10.1088/1742-6596/363/1/012038.

[137] O.J. Piccinni et al. A new data analysis framework for the search of continuous gravitational wave signals. *Classical and Quantum Gravity*, 36(1):015008, dec 2018. doi: 10.1088/1361-6382/aaefb5. URL https://dx.doi.org/10.1088/1361-6382/aaefb5.

[138] G. Poghosyan et al. Architecture, implementation and parallelization of the software to search for periodic gravitational wave signals. *Computer Physics Communications*, 188:167–176, 2015. ISSN 0010-4655. doi: https://doi.org/10.1016/j.cpc.2014.10.025. URL https://www.sciencedirect.com/science/article/pii/S0010465514003774.

[139] LIGO Scientific Collaboration. LIGO Algorithm Library - LALSuite, 2018. Free software (GPL).

[140] P. Patel et al. Implementation of barycentric resampling for continuous wave searches in gravitational wave data. *Phys. Rev. D*, 81:084032, Apr 2010. doi: 10.1103/PhysRevD.81.084032. URL https://link.aps.org/doi/10.1103/PhysRevD.81.084032.

[141] R. Prix. Characterizing timing and memory-requirements of the f-statistic implementations in lalsuite. Tech. Rep. T1600531-v4, LIGO, 2017. URL https://dcc.ligo.org/LIGO-T1600531-v4/public.

[142] P. R. Williams and B. F. Schutz. An efficient matched filtering algorithm for the detection of continuous gravitational wave signals. *AIP Conference Proceedings*, 523(1):473–476, 06 2000. ISSN 0094-243X. doi: 10.1063/1.1291918. URL https://doi.org/10.1063/1.1291918.

[143] S. Suvorova et al. Hidden markov model tracking of continuous gravitational waves from a binary neutron star with wandering spin. ii. binary orbital phase tracking. *Phys. Rev. D*, 96:102006, Nov 2017. doi: 10.1103/PhysRevD.96.102006. URL https://link.aps.org/doi/10.1103/PhysRevD.96.102006.

[144] L. Sammut, C. Messenger, A. Melatos, and B. J. Owen. Implementation of the frequency-modulated sideband search method for gravitational waves from low mass x-ray binaries. *Phys. Rev. D*, 89:043001, Feb 2014. doi: 10.1103/PhysRevD.89.043001. URL https://link.aps.org/doi/10.1103/PhysRevD.89.043001.

[145] A. C. Searle. Monte-Carlo and Bayesian techniques in gravitational wave burst data analysis. In *12th Gravitational Wave Data Analysis Workshop*, 4 2008.

[146] R. Prix and B. Krishnan. Targeted search for continuous gravitational waves: Bayesian versus maximum-likelihood statistics. *Classical and Quantum Gravity*, 26(20):204013, oct 2009. doi: 10.1088/0264-9381/26/20/204013. URL https://dx.doi.org/10.1088/0264-9381/26/20/204013.

[147] R. T. Edwards, G. B. Hobbs, and R. N. Manchester. tempo2, a new pulsar timing package – II. The timing model and precision estimates. *Monthly Notices of the Royal Astronomical Society*,

372(4):1549–1574, 10 2006. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2006.10870.x. URL https://doi.org/10.1111/j.1365-2966.2006.10870.x.

[148] P. Pitkin and G. Woan. Searching for gravitational waves from the crab pulsar—the problem of timing noise. *Classical and Quantum Gravity*, 21(5):S843, feb 2004. doi: 10.1088/0264-9381/21/5/069. URL https://dx.doi.org/10.1088/0264-9381/21/5/069.

[149] Réjean J. Dupuis and Graham Woan. Bayesian estimation of pulsar parameters from gravitational wave data. *Phys. Rev. D*, 72:102002, Nov 2005. doi: 10.1103/PhysRevD.72.102002. URL https://link.aps.org/doi/10.1103/PhysRevD.72.102002.

[150] M. Pitkin et al. A nested sampling code for targeted searches for continuous gravitational waves from pulsars, 2017. URL https://arxiv.org/abs/1705.08978.

[151] M. Pitkin et al. First results and future prospects for dual-harmonic searches for gravitational waves from spinning neutron stars. *Monthly Notices of the Royal Astronomical Society*, 453(4):4399–4420, 09 2015. ISSN 0035-8711. doi: 10.1093/mnras/stv1931. URL https://doi.org/10.1093/mnras/stv1931.

[152] M. Isi, M. Pitkin, and A. J. Weinstein. Probing dynamical gravity with the polarization of continuous gravitational waves. *Phys. Rev. D*, 96:042001, Aug 2017. doi: 10.1103/PhysRevD.96.042001. URL https://link.aps.org/doi/10.1103/PhysRevD.96.042001.

[153] Matthew Pitkin. Cwinpy: A python package for inference with continuous gravitational-wave signals from pulsars. *Journal of Open Source Software*, 7(77):4568, 2022. doi: 10.21105/joss.04568. URL https://doi.org/10.21105/joss.04568.

[154] G. Ashton et al. Bilby: A user-friendly bayesian inference library for gravitational-wave astronomy. *The Astrophysical Journal Supplement Series*, 241(2):27, apr 2019. doi: 10.3847/1538-4365/ab06fc. URL https://dx.doi.org/10.3847/1538-4365/ab06fc.

[155] B. P. Abbott et al. Multi-messenger observations of a binary neutron star merger*. *The Astrophysical Journal Letters*, 848(2):L12, oct 2017. doi: 10.3847/2041-8213/aa91c9. URL https://dx.doi.org/10.3847/2041-8213/aa91c9.

[156] B. Abbott et al. Einstein@home search for periodic gravitational waves in ligo s4 data. *Phys. Rev. D*, 79:022001, Jan 2009. doi: 10.1103/PhysRevD.79.022001. URL https://link.aps.org/doi/10.1103/PhysRevD.79.022001.

[157] P. V. C. Hough. Machine Analysis of Bubble Chamber Pictures. *Conf. Proc. C*, 590914:554–558, 1959.

[158] B. Krishnan et al. Hough transform search for continuous gravitational waves. *Phys. Rev. D*, 70:082001, Oct 2004. doi: 10.1103/PhysRevD.70.082001. URL https://link.aps.org/doi/10.1103/PhysRevD.70.082001.

[159] F. Antonucci et al. Detection of periodic gravitational wave sources by hough transform in the f versus plane. *Classical and Quantum Gravity*, 25(18):184015, sep 2008. doi: 10.1088/0264-9381/25/18/184015. URL https://dx.doi.org/10.1088/0264-9381/25/18/184015.

[160] C. Palomba, P. Astone, and S. Frasca. Adaptive hough transform for the search of periodic sources. *Classical and Quantum Gravity*, 22(18):S1255, sep 2005. doi: 10.1088/0264-9381/22/18/S39. URL https://dx.doi.org/10.1088/0264-9381/22/18/S39.

[161] P. B. Covas and Alicia M. Sintes. New method to search for continuous gravitational waves from unknown neutron stars in binary systems. *Phys. Rev. D*, 99:124019, Jun 2019. doi: 10.1103/PhysRevD.99.124019. URL https://link.aps.org/doi/10.1103/PhysRevD.99.124019.

[162] O. J. Piccinni et al. Directed search for continuous gravitational-wave signals from the galactic center in the advanced ligo second observing run. *Phys. Rev. D*, 101:082004, Apr 2020. doi: 10.1103/PhysRevD.101.082004. URL https://link.aps.org/doi/10.1103/PhysRevD.101.082004.

[163] P. Leaci et al. Novel directed search strategy to detect continuous gravitational waves from neutron stars in low- and high-eccentricity binary systems. *Phys. Rev. D*, 95:122001, Jun 2017. doi: 10.1103/PhysRevD.95.122001. URL https://link.aps.org/doi/10.1103/PhysRevD.95.122001.

[164] A. Miller et al. Method to search for long duration gravitational wave transients from isolated neutron stars using the generalized frequency-hough transform. *Phys. Rev. D*, 98:102004, Nov 2018. doi: 10.1103/PhysRevD.98.102004. URL https://link.aps.org/doi/10.1103/PhysRevD.98.102004.

[165] G. Mendell and M. Landry. Stackslide and hough search snr and statistics. Tech. Rep. T050003-x0, LIGO, 2005. URL https://dcc.ligo.org/LIGO-T050003-x0/public.

[166] P. Jaranowski and A. Królak. Gravitational-wave data analysis. formalism and sample applications: The gaussian case. *Living Reviews in Relativity*, 8(1):3, Mar 2005. ISSN 1433-8351. doi: 10.12942/lrr-2005-3. URL https://doi.org/10.12942/lrr-2005-3.

[167] B. Allen. Optimal template banks. *Phys. Rev. D*, 104:042005, Aug 2021. doi: 10.1103/PhysRevD.104.042005. URL https://link.aps.org/doi/10.1103/PhysRevD.104.042005.

[168] K. Wette. Empirically extending the range of validity of parameter-space metrics for all-sky searches for gravitational-wave pulsars. *Phys. Rev. D*, 94:122002, Dec 2016. doi: 10.1103/PhysRevD.94.122002. URL https://link.aps.org/doi/10.1103/PhysRevD.94.122002.

[169] H. J. Pletsch and B. Allen. Exploiting large-scale correlations to detect continuous gravitational waves. *Phys. Rev. Lett.*, 103:181102, Oct 2009. doi: 10.1103/PhysRevLett.103.181102. URL https://link.aps.org/doi/10.1103/PhysRevLett.103.181102.

[170] K. Wette. Lattice template placement for coherent all-sky searches for gravitational-wave pulsars. *Phys. Rev. D*, 90:122010, Dec 2014. doi: 10.1103/PhysRevD.90.122010. URL https://link.aps.org/doi/10.1103/PhysRevD.90.122010.

[171] K. Wette, S. Walsh, R. Prix, and M. A. Papa. Implementing a semicoherent search for continuous gravitational waves using optimally constructed template banks. *Phys. Rev. D*, 97:123016, Jun 2018. doi: 10.1103/PhysRevD.97.123016. URL https://link.aps.org/doi/10.1103/PhysRevD.97.123016.

[172] S. Walsh, K. Wette, M. A. Papa, and R. Prix. Optimizing the choice of analysis method for all-sky searches for continuous gravitational waves with einstein@home. *Phys. Rev. D*, 99:082004, Apr 2019. doi: 10.1103/PhysRevD.99.082004. URL https://link.aps.org/doi/10.1103/PhysRevD.99.082004.

[173] V. Dergachev. Description of powerflux algorithms and implementation. Tech. Rep. T050186-x0, LIGO, 2005. URL https://dcc.ligo.org/LIGO-T050186-x0/public.

[174] V. Dergachev. Novel universal statistic for computing upper limits in an ill-behaved background. *Phys. Rev. D*, 87:062001, Mar 2013. doi: 10.1103/PhysRevD.87.062001. URL https://link.aps.org/doi/10.1103/PhysRevD.87.062001.

[175] V. Dergachev. On blind searches for noise dominated signals: a loosely coherent approach. *Classical and Quantum Gravity*, 27(20):205017, sep 2010. doi: 10.1088/0264-9381/27/20/205017. URL https://dx.doi.org/10.1088/0264-9381/27/20/205017.

[176] Holger J. Pletsch. Sliding coherence window technique for hierarchical detection of continuous gravitational waves. *Phys. Rev. D*, 83:122003, Jun 2011. doi: 10.1103/PhysRevD.83.122003. URL https://link.aps.org/doi/10.1103/PhysRevD.83.122003.

[177] V. Dergachev and M. A. Papa. Sensitivity improvements in the search for periodic gravitational waves using o1 ligo data. *Phys. Rev. Lett.*, 123:101101, Sep 2019. doi: 10.1103/PhysRevLett.123.101101. URL https://link.aps.org/doi/10.1103/PhysRevLett.123.101101.

[178] S. Ballmer. A radiometer for stochastic gravitational waves. *Classical and Quantum Gravity*, 23(8):S179, mar 2006. doi: 10.1088/0264-9381/23/8/S23. URL https://dx.doi.org/10.1088/0264-9381/23/8/S23.

[179] S. Dhurandhar et al. Cross-correlation search for periodic gravitational waves. *Phys. Rev. D*, 77:082001, Apr 2008. doi: 10.1103/PhysRevD.77.082001. URL https://link.aps.org/doi/10.1103/PhysRevD.77.082001.

[180] C. T. Y. Chung, A. Melatos, B. Krishnan, and J. T. Whelan. Designing a cross-correlation search for continuous-wave gravitational radiation from a neutron star in the supernova remnant SNR 1987A*. *Monthly Notices of the Royal Astronomical Society*, 414(3):2650–2663, 06 2011. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2011.18585.x. URL https://doi.org/10.1111/j.1365-2966.2011.18585.x.

[181] K. Wagner et al. Template lattices for a cross-correlation search for gravitational waves from scorpius x-1. *Classical and Quantum Gravity*, 39(7):075013, mar 2022. doi: 10.1088/1361-6382/ac5012. URL https://dx.doi.org/10.1088/1361-6382/ac5012.

[182] E. Goetz and K. Riles. An all-sky search algorithm for continuous gravitational waves from spinning neutron stars in binary systems. *Classical and Quantum Gravity*, 28(21):215006, sep 2011. doi: 10.1088/0264-9381/28/21/215006. URL https://dx.doi.org/10.1088/0264-9381/28/21/215006.

[183] G. Meadors, E. Goetz, and K. Riles. Tuning into scorpius x-1: adapting a continuous gravitational-wave search for a known binary system. *Classical and Quantum Gravity*, 33(10):105017, apr 2016. doi: 10.1088/0264-9381/33/10/105017. URL https://dx.doi.org/10.1088/0264-9381/33/10/105017.

[184] Joe Bayley, Chris Messenger, and Graham Woan. Generalized application of the viterbi algorithm to searches for continuous gravitational-wave signals. *Phys. Rev. D*, 100:023006, Jul 2019. doi: 10.1103/PhysRevD.100.023006. URL https://link.aps.org/doi/10.1103/PhysRevD.100.023006.

[185] L. Sun, A. Melatos, S. Suvorova, W. Moran, and R. J. Evans. Hidden markov model tracking of continuous gravitational waves from young supernova remnants. *Phys. Rev. D*, 97:043013, Feb 2018. doi: 10.1103/PhysRevD.97.043013. URL https://link.aps.org/doi/10.1103/PhysRevD.97.043013.

[186] J. Bayley, C. Messenger, and G. Woan. Robust machine learning algorithm to search for continuous gravitational waves. *Phys. Rev. D*, 102:083024, Oct 2020. doi: 10.1103/PhysRevD.102.083024. URL https://link.aps.org/doi/10.1103/PhysRevD.102.083024.

[187] C. Cutler, I. Gholami, and B. Krishnan. Improved stack-slide searches for gravitational-wave pulsars. *Phys. Rev. D*, 72:042004, Aug 2005. doi: 10.1103/PhysRevD.72.042004. URL https://link.aps.org/doi/10.1103/PhysRevD.72.042004.

[188] C. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, 1995.

[189] G. E. Hinton, S. Osindero, and Y. Teh. A fast learning algorithm for deep belief nets. *Neural Comput.*, 18(7):1527–1554, jul 2006. ISSN 0899-7667. doi: 10.1162/neco.2006.18.7.1527. URL https://doi.org/10.1162/neco.2006.18.7.1527.

[190] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'12, page 1097–1105, Red Hook, NY, USA, 2012. Curran Associates Inc.

[191] 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)*, pages 770–778, 2016. doi: 10.1109/CVPR.2016.90.

[192] François Chollet et al. Keras. https://keras.io, 2015.

[193] J. Ansel et al. PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation. In *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)*. ACM, April 2024. doi: 10.1145/3620665.3640366. URL https://pytorch.org/assets/pytorch2-2.pdf.

[194] M. Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL https://www.tensorflow.org/. Software available from tensorflow.org.

[195] V. Nair and G. E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on Machine Learning*, pages 807–814, Madison, WI, USA, 2010. Omnipress.

[196] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989. ISSN 0893-6080. doi: https://doi.org/10.1016/0893-6080(89)90020-8. URL https://www.sciencedirect.com/science/article/pii/0893608089900208.

[197] C. Zhang et al. Understanding deep learning requires rethinking generalization, 2017. URL https://arxiv.org/abs/1611.03530.

[198] M. Advani and A. Saxe. High-dimensional dynamics of generalization error in neural networks, 2017. URL https://arxiv.org/abs/1710.03667.

[199] M.A. Nielsen. *Neural Networks and Deep Learning*. Determination Press, 2015. URL https://books.google.co.jp/books?id=STDBswEACAAJ.

[200] D. Rumelhart and D. Zipser. Feature discovery by competitive learning. *Cognitive Science*, 9 (1):75–112, 1985. ISSN 0364-0213. doi: https://doi.org/10.1016/S0364-0213(85)80010-0. URL https://www.sciencedirect.com/science/article/pii/S0364021385800100.

[201] B. Neal et al. A modern take on the bias-variance tradeoff in neural networks, 2019. URL https://arxiv.org/abs/1810.08591.

[202] Andrew Y. Ng. Feature selection, l1 vs. l2 regularization, and rotational invariance. In *Proceedings of the Twenty-First International Conference on Machine Learning*, ICML '04, page 78, New York, NY, USA, 2004. Association for Computing Machinery. ISBN 1581138385. doi: 10.1145/1015330.1015435. URL https://doi.org/10.1145/1015330.1015435.

[203] A. Wilson et al. The marginal value of adaptive gradient methods in machine learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pages 4151–4161, 2017.

[204] L. Bottou. *Stochastic Gradient Descent Tricks*, pages 421–436. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-35289-8. doi: 10.1007/978-3-642-35289-8_25. URL https://doi.org/10.1007/978-3-642-35289-8_25.

[205] D. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–536, 1986. URL https://api.semanticscholar.org/CorpusID:205001834.

[206] C. M. Bishop. Training with noise is equivalent to tikhonov regularization. *Neural Computation*, 7(1):108–116, 1995. doi: 10.1162/neco.1995.7.1.108.

[207] N. Keskar et al. On large-batch training for deep learning: Generalization gap and sharp minima. *CoRR*, abs/1609.04836, 2016. URL http://arxiv.org/abs/1609.04836.

[208] Y. Nesterov. A method of solving a convex programming problem with convergence rate o(1/k^2). *Soviet Mathematics Doklady*, 27 : 372 − −376, 1983.

[209] N. Qian. On the momentum term in gradient descent learning algorithms. *Neural Networks*, 12(1): 145–151, 1999. ISSN 0893-6080. doi: https://doi.org/10.1016/S0893-6080(98)00116-6. URL https://www.sciencedirect.com/science/article/pii/S0893608098001166.

[210] 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 International Conference on Machine Learning - Volume 28*, ICML'13, page III–1139–III–1147. JMLR.org, 2013.

[211] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(61):2121–2159, 2011. URL http://jmlr.org/papers/v12/duchi11a.html.

[212] A. Graves. Generating sequences with recurrent neural networks, 2014. URL https://arxiv.org/abs/1308.0850.

[213] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017. URL https://arxiv.org/abs/1412.6980.

[214] N. Srivastava et al. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958, jan 2014. ISSN 1532-4435.

[215] S. Ioffe and C. Szegedy. Batch normalization: accelerating deep network training by reducing internal covariate shift. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*, ICML'15, page 448–456. JMLR.org, 2015.

[216] Li Deng. The mnist database of handwritten digit images for machine learning research. *IEEE Signal Processing Magazine*, 29(6):141–142, 2012.

[217] Y. Lecun and Y. Bengio. *Convolutional networks for images, speech, and time-series*. MIT Press, 1995.

[218] E Miles Stoudenmire. Learning relevant features of data with multi-scale tensor networks. *Quantum Science and Technology*, 3(3):034003, 2018. ISSN 2058-9565. doi: 10.1088/2058-9565/aaba1a. URL http://dx.doi.org/10.1088/2058-9565/aaba1a.

[219] Y. Levine, D. Yakira, N. Cohen, and A. Shashua. Deep learning and quantum entanglement: Fundamental connections with implications to network design, 2017. URL https://arxiv.org/abs/1704.01552.

[220] Taco S. Cohen and Max Welling. Group equivariant convolutional networks, 2016. URL https://arxiv.org/abs/1602.07576.

[221] C. Szegedy et al. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015. doi: 10.1109/CVPR.2015.7298594.

[222] C. Szegedy et al. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016. doi: 10.1109/CVPR.2016.308.

[223] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition, 2015. URL https://arxiv.org/abs/1409.1556.

[224] Y. Jia et al. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.

[225] Reinhard Prix. *Gravitational Waves from Spinning Neutron Stars*, pages 651–685. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. ISBN 978-3-540-76965-1. doi: 10.1007/978-3-540-76965-1_24. URL https://doi.org/10.1007/978-3-540-76965-1_24.

[226] Magdalena Sieniawska and Michał Bejger. Continuous gravitational waves from neutron stars: Current status and prospects. *Universe*, 5(11):217, 2019. ISSN 2218-1997. doi: 10.3390/universe5110217. URL https://www.mdpi.com/2218-1997/5/11/217.

[227] B. P. Abbott et al. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 116:061102, Feb 2016. doi: 10.1103/PhysRevLett.116.061102. URL https://link.aps.org/doi/10.1103/PhysRevLett.116.061102.

[228] B. P. Abbott et al. GW170817: Observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett.*, 119:161101, Oct 2017. doi: 10.1103/PhysRevLett.119.161101. URL https://link.aps.org/doi/10.1103/PhysRevLett.119.161101.

[229] B. P. Abbott et al. GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs. *Phys. Rev. X*, 9:031040, Sep 2019. doi: 10.1103/PhysRevX.9.031040. URL https://link.aps.org/doi/10.1103/PhysRevX.9.031040.

[230] R. Abbott et al. GWTC-2: Compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run. *Phys. Rev. X*, 11:021053, Jun 2021. doi: 10.1103/PhysRevX.11.021053. URL https://link.aps.org/doi/10.1103/PhysRevX.11.021053.

[231] R. Abbott et al. GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run. *Phys. Rev. X*, 13:041039, Dec 2023. doi: 10.1103/PhysRevX.13.041039. URL https://link.aps.org/doi/10.1103/PhysRevX.13.041039.

[232] Keith Riles. Recent searches for continuous gravitational waves. *Modern Physics Letters A*, 32(39):1730035, 2017. doi: 10.1142/S021773231730035X. URL https://doi.org/10.1142/S021773231730035X.

[233] Yusifov, I. and Küçük, I. Revisiting the radial distribution of pulsars in the galaxy. *A & A*, 422(2):545–553, 2004. doi: 10.1051/0004-6361:20040152. URL https://doi.org/10.1051/0004-6361:20040152.

[234] A. G. Lyne, R. N. Manchester, and J. H. Taylor. The galactic population of pulsars. *Monthly Notices of the Royal Astronomical Society*, 213(3):613–639, 04 1985. ISSN 0035-8711. doi: 10.1093/mnras/213.3.613. URL https://doi.org/10.1093/mnras/213.3.613.

[235] B. P. Abbott et al. Prospects for observing and localizing gravitational-wave transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Reviews in Relativity*, 23(1):3, 2020. doi: 10.1007/s41114-020-00026-9. URL https://doi.org/10.1007/s41114-020-00026-9.

[236] Paul D. Lasky. Gravitational waves from neutron stars: A review. *Publications of the Astronomical Society of Australia*, 32:e034, 2015. doi: 10.1017/pasa.2015.35.

[237] Piotr Jaranowski and Andrzej Królak. Data analysis of gravitational-wave signals from spinning neutron stars. iii. detection statistics and computational requirements. *Phys. Rev. D*, 61:062001, Feb 2000. doi: 10.1103/PhysRevD.61.062001. URL https://link.aps.org/doi/10.1103/PhysRevD.61.062001.

[238] R. Prix and M. Shaltev. Search for continuous gravitational waves: Optimal stackslide method at fixed computing cost. *Phys. Rev. D*, 85:084010, Apr 2012. doi: 10.1103/PhysRevD.85.084010. URL https://link.aps.org/doi/10.1103/PhysRevD.85.084010.

[239] Lenka Zdeborová. Understanding deep learning is also a job for physicists. *Nature Physics*, 16(6):602–604, Jun 2020. ISSN 1745-2481. doi: 10.1038/s41567-020-0929-2. URL https://doi.org/10.1038/s41567-020-0929-2.

[240] Daniel A. Roberts, Sho Yaida, and Boris Hanin. *Neural Networks*, page 37–52. Cambridge University Press, 2022. doi: 10.1017/9781009023405.004.

[241] E. Cuoco et al. Enhancing gravitational-wave science with machine learning. *Machine Learning: Science and Technology*, 2(1):011002, dec 2020. doi: 10.1088/2632-2153/abb93a. URL https://dx.doi.org/10.1088/2632-2153/abb93a.

[242] Daniel George and E.A. Huerta. Deep learning for real-time gravitational wave detection and parameter estimation: Results with Advanced LIGO data. *Physics Letters B*, 778:64–70, 2018. ISSN 0370-2693. doi: https://doi.org/10.1016/j.physletb.2017.12.053. URL https://www.sciencedirect.com/science/article/pii/S0370269317310390.

[243] Hongyu Shen, E A Huerta, Eamonn O'Shea, Prayush Kumar, and Zhizhen Zhao. Statistically-informed deep learning for gravitational wave parameter estimation. *Machine Learning: Science and Technology*, 3 (1):015007, nov 2021. doi: 10.1088/2632-2153/ac3843. URL https://dx.doi.org/10.1088/2632-2153/ac3843.

[244] G. Vajente et al. Machine-learning nonstationary noise out of gravitational-wave detectors. *Phys. Rev. D*, 101:042003, Feb 2020. doi: 10.1103/PhysRevD.101.042003. URL https://link.aps.org/doi/10.1103/PhysRevD.101.042003.

[245] Zoheyr Doctor, Ben Farr, Daniel E. Holz, and Michael Pürrer. Statistical gravitational waveform models: What to simulate next? *Phys. Rev. D*, 96:123011, Dec 2017. doi: 10.1103/PhysRevD.96.123011. URL https://link.aps.org/doi/10.1103/PhysRevD.96.123011.

[246] Yoshinta Setyawati, Michael Pürrer, and Frank Ohme. Regression methods in waveform modeling: a comparative study. *Classical and Quantum Gravity*, 37(7):075012, mar 2020. doi: 10.1088/1361-6382/ab693b. URL https://dx.doi.org/10.1088/1361-6382/ab693b.

[247] J. Powell et al. Classification methods for noise transients in advanced gravitational-wave detectors II: performance tests on Advanced LIGO data. *Classical and Quantum Gravity*, 34(3):034002, jan 2017. doi: 10.1088/1361-6382/34/3/034002. URL https://dx.doi.org/10.1088/1361-6382/34/3/034002.

[248] M Cavaglià et al. Improving the background of gravitational-wave searches for core collapse supernovae: a machine learning approach. *Machine Learning: Science and Technology*, 1(1):015005, feb 2020. doi: 10.1088/2632-2153/ab527d. URL https://dx.doi.org/10.1088/2632-2153/ab527d.

[249] Christoph Dreissigacker and Reinhard Prix. Deep-learning continuous gravitational waves: Multiple detectors and realistic noise. *Phys. Rev. D*, 102:022005, Jul 2020. doi: 10.1103/PhysRevD.102.022005. URL https://link.aps.org/doi/10.1103/PhysRevD.102.022005.

[250] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi. Inception-v4, inception-ResNet and the impact of residual connections on learning. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 4278–4284. AAAI Press, 2017.

[251] Joe Bayley, Chris Messenger, and Graham Woan. Robust machine learning algorithm to search for continuous gravitational waves. *Phys. Rev. D*, 102:083024, Oct 2020. doi: 10.1103/PhysRevD.102.083024. URL https://link.aps.org/doi/10.1103/PhysRevD.102.083024.

[252] Joe Bayley, Chris Messenger, and Graham Woan. Generalized application of the viterbi algorithm to searches for continuous gravitational-wave signals. *Phys. Rev. D*, 100:023006, Jul 2019. doi: 10.1103/PhysRevD.100.023006. URL https://link.aps.org/doi/10.1103/PhysRevD.100.023006.

[253] Prasanna M. Joshi and Reinhard Prix. Novel neural-network architecture for continuous gravitational waves. *Phys. Rev. D*, 108:063021, Sep 2023. doi: 10.1103/PhysRevD.108.063021. URL https://link.aps.org/doi/10.1103/PhysRevD.108.063021.

[254] Luana M. Modafferi, Rodrigo Tenorio, and David Keitel. Convolutional neural network search for long-duration transient gravitational waves from glitching pulsars. *Phys. Rev. D*, 108:023005, Jul 2023. doi: 10.1103/PhysRevD.108.023005. URL https://link.aps.org/doi/10.1103/PhysRevD.108.023005.

[255] Jason Yosinski, Jeff Clune, Y. Bengio, and Hod Lipson. How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems (NIPS)*, 27, 11 2014.

[256] Daniel George, Hongyu Shen, and E. A. Huerta. Classification and unsupervised clustering of ligo data with deep transfer learning. *Phys. Rev. D*, 97:101501, May 2018. doi: 10.1103/PhysRevD.97.101501. URL https://link.aps.org/doi/10.1103/PhysRevD.97.101501.

[257] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He. Aggregated residual transformations for deep neural networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5987–5995, Los Alamitos, CA, USA, jul 2017. IEEE Computer Society. doi: 10.1109/CVPR.2017.634. URL https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.634.

[258] Keith Riles. Searches for continuous-wave gravitational radiation. *Living Reviews in Relativity*, 26(1): 3, Apr 2023. ISSN 1433-8351. doi: 10.1007/s41114-023-00044-3. URL https://doi.org/10.1007/s41114-023-00044-3.

[259] Piotr Jaranowski, Andrzej Królak, and Bernard F. Schutz. Data analysis of gravitational-wave signals from spinning neutron stars: The signal and its detection. *Phys. Rev. D*, 58:063001, Aug 1998. doi: 10.1103/PhysRevD.58.063001. URL https://link.aps.org/doi/10.1103/PhysRevD.58.063001.

[260] S. Suvorova, L. Sun, A. Melatos, W. Moran, and R. J. Evans. Hidden markov model tracking of continuous gravitational waves from a neutron star with wandering spin. *Phys. Rev. D*, 93:123009, Jun 2016. doi: 10.1103/PhysRevD.93.123009. URL https://link.aps.org/doi/10.1103/PhysRevD.93.123009.

[261] Fuzhen Zhuang et al. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1): 43–76, 2021. doi: 10.1109/JPROC.2020.3004555.

[262] A. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. Cnn features off-the-shelf: An astounding baseline for recognition. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 512–519, Los Alamitos, CA, USA, jun 2014. IEEE Computer Society. doi: 10.1109/CVPRW.2014.131. URL https://doi.ieeecomputersociety.org/10.1109/CVPRW.2014.131.

[263] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity mappings in deep residual networks. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 630–645, Cham, 2016. Springer International Publishing. ISBN 978-3-319-46493-0.

[264] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, Las Vegas, NV, USA, 2016. IEEE. doi: 10.1109/CVPR.2016.90.

[265] Adam Paszke et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[266] R. Abbott et al. Open data from the third observing run of LIGO, Virgo, KAGRA, and GEO. *The Astrophysical Journal Supplement Series*, 267(2):29, jul 2023. doi: 10.3847/1538-4365/acdc9f. URL https://dx.doi.org/10.3847/1538-4365/acdc9f.

[267] David Keitel, Rodrigo Tenorio, Gregory Ashton, and Reinhard Prix. PyFstat: a Python package for continuous gravitational-wave data analysis. *J. Open Source Softw.*, 6(60):3000, 2021. doi: 10.21105/joss.03000.

[268] Gregory Ashton, David Keitel, Reinhard Prix, and Rodrigo Tenorio. PyFstat, July 2020. URL https://doi.org/10.5281/zenodo.3967045. 10.5281/zenodo.3967045.

[269] Piotr Jaranowski and Andrzej Królak. Data analysis of gravitational-wave signals from spinning neutron stars. iii. detection statistics and computational requirements. *Phys. Rev. D*, 61:062001, Feb 2000. doi: 10.1103/PhysRevD.61.062001. URL https://link.aps.org/doi/10.1103/PhysRevD.61.062001.

[270] Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, page 41–48,

New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605585161. doi: 10.1145/1553374.1553380. URL https://doi.org/10.1145/1553374.1553380.

[271] Petru Soviany, Radu Tudor Ionescu, Paolo Rota, and Nicu Sebe. Curriculum learning: A survey. *International Journal of Computer Vision*, 130(6):1526–1565, Jun 2022. ISSN 1573-1405. doi: 10.1007/s11263-022-01611-x. URL https://doi.org/10.1007/s11263-022-01611-x.

[272] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[273] Diederik Kingman and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.

[274] Zhilu Zhang and Mert R. Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, page 8792–8802, Montréal, Canada, 2018. Curran Associates Inc.

[275] R. Abbott et al. Search of the early o3 ligo data for continuous gravitational waves from the cassiopeia a and vela jr. supernova remnants. *Phys. Rev. D*, 105:082005, Apr 2022. doi: 10.1103/PhysRevD.105.082005. URL https://link.aps.org/doi/10.1103/PhysRevD.105.082005.

[276] K. Wette, S. Walsh, R. Prix, and M. A. Papa. Implementing a semicoherent search for continuous gravitational waves using optimally constructed template banks. *Phys. Rev. D*, 97:123016, Jun 2018. doi: 10.1103/PhysRevD.97.123016. URL https://link.aps.org/doi/10.1103/PhysRevD.97.123016.

[277] R. Abbott et al. Search of the early o3 ligo data for continuous gravitational waves from the cassiopeia a and vela jr. supernova remnants. *Phys. Rev. D*, 105:082005, Apr 2022. doi: 10.1103/PhysRevD.105.082005. URL https://link.aps.org/doi/10.1103/PhysRevD.105.082005.

[278] B. Chandra and Manish Gupta. Robust approach for estimating probabilities in naïve–bayes classifier for gene expression data. *Expert Systems with Applications*, 38(3):1293–1298, 2011. ISSN 0957-4174. doi: https://doi.org/10.1016/j.eswa.2010.06.076. URL https://www.sciencedirect.com/science/article/pii/S0957417410005804.

[279] B. S. Sathyaprakash. Filtering post-newtonian gravitational waves from coalescing binaries. *Phys. Rev. D*, 50:R7111–R7115, Dec 1994. doi: 10.1103/PhysRevD.50.R7111. URL https://link.aps.org/doi/10.1103/PhysRevD.50.R7111.

[280] R. Balasubramanian and S. V. Dhurandhar. Performance of newtonian filters in detecting gravitational waves from coalescing binaries. *Phys. Rev. D*, 50:6080–6088, Nov 1994. doi: 10.1103/PhysRevD.50.6080. URL https://link.aps.org/doi/10.1103/PhysRevD.50.6080.

[281] D. C. Backer and R. W. Hellings. Pulsar timing and general relativity. *Annual Rev. Astron. Astrophys.*, 24:537–575, January 1986. doi: 10.1146/annurev.aa.24.090186.002541.

[282] V. Szebehely. *Applications of celestial mechanics for deep space investigations*. doi: 10.2514/6.1968-1123. URL https://arc.aiaa.org/doi/abs/10.2514/6.1968-1123.

[283] R. Blandford and S. A. Teukolsky. Arrival-time analysis for a pulsar in a binary system. *Astrophysical Journal*, 205:580–591, April 1976. doi: 10.1086/154315.

[284] Irwin I. Shapiro. Fourth Test of General Relativity. *Physical Review Letters*, 13(26):789–791, December 1964. doi: 10.1103/PhysRevLett.13.789.