

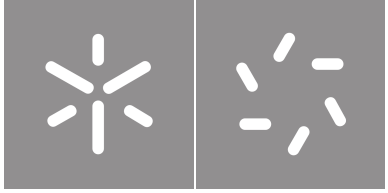
Universidade do Minho
Escola de Ciências

Solange da Silva Nunes

Deep-Learning Inference of Rotational Core-Collapse Supernovae with Numerically-Generated Gravitational-Wave Signals

April, 2023

Deep-Learning Inference of Rotational Core-Collapse Supernovae with Numerically-Generated Gravitational-Wave Signals
Solange Nunes
UMinho | 2023



Universidade do Minho
Escola de Ciências

Solange da Silva Nunes

Deep-Learning Inference of Rotational Core-Collapse Supernovae with Numerically-Generated Gravitational-Wave Signals

Master Thesis
Master in Physics

Work developed under the supervision of:
Prof. Dr. António Joaquim Onofre Abreu Ribeiro Gonçalves
Prof. Dr. Alejandro Torres-Forné

April, 2023

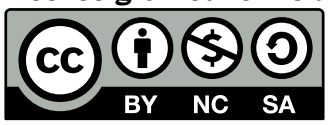
COPYRIGHT AND TERMS OF USE OF THIS WORK BY A THIRD PARTY

This is academic work that can be used by third parties as long as internationally accepted rules and good practices regarding copyright and related rights are respected.

Accordingly, this work may be used under the license provided below.

If the user needs permission to make use of the work under conditions not provided for in the indicated licensing, they should contact the author through the RepositoriUM of Universidade do Minho.

License granted to the users of this work



Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International

CC BY-NC-SA 4.0

https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en

This document was created with the (pdf/Xe/Lua)LaTeX processor and the NOVAthesis template (v7.0.0) [1].

Acknowledgements

I would like to take this opportunity to express my gratitude to the individuals who have supported me throughout my academic journey. I am deeply thankful for the guidance, expertise, and encouragement that I have received from my advisors, colleagues, family, and friends.

Firstly, I would like to express my deepest appreciation to my advisors, Professor António Onofre and Professor Alejandro Torres-Forné. Professor António Onofre has been a constant source of inspiration for me. His passion and enthusiasm have been fundamental in shaping my academic journey. Professor Alejandro Torres-Forné has also been instrumental in my academic growth, and I am grateful for his expert guidance and support. I feel honored to have them both as my advisors.

I cannot emphasize enough the contribution of Professor José Font, whose insights and encouragement deeply influenced my work. I feel fortunate to have the opportunity to work with him.

I would also like to thank Osvaldo for his support and availability in helping whenever issues arose.

Finally, I want to express my sincere gratitude to Manuel, who has been a constant source of support, encouragement and motivation. His unwavering belief in me has been invaluable, and I consider myself extremely lucky to have had him by my side on every step of the way.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the Universidade do Minho.

Braga, 30 April 2023

(Place) (Date)

Assinado por: **SOLANGE DA SILVA NUNES**
Num. de Identificação: 14165553
Data: 2023.04.30 22:30:20+01'00'

(Solange da Silva Nunes)

Resumo

Inferência com Aprendizagem Profunda de Ondas Gravitacionais de Supernovas de Colapso Nuclear em Rotação Geradas Numericamente

O colapso gravitacional do núcleo de estrelas massivas e a sua subsequente explosão como Supernovas (SN) é uma das fontes de ondas gravitacionais (GW) mais interessantes que esperamos detetar nos próximos anos. As supernovas de colapso nuclear (CCSNe) apresentam um grande interesse científico pois a análise das suas complexas ondas pode potencialmente providenciar informação valiosa sobre os processos físicos que ocorrem durante o colapso gravitacional dos núcleos de ferro de estrelas massivas. Neste trabalho, usam-se técnicas de *Aprendizagem Profunda* para analisar sinais de GW de CCSN em rotação, classificando-as e separando-as do ruído de fundo dos detetores e estimando alguns parâmetros físicos dos sistemas que produzem estes sinais.

Para detetar GW de CCSN em rotação em dados com ruído de detetores, treinou-se uma rede de classificação com um conjunto de dados de *Series Temporais* de GW de CCSN geradas numericamente e injetadas em ruído real dos detetores Advanced LIGO e Advanced Virgo (correspondente ao período de observação O3a). O teste de classificação alcançou resultados significativos, atingindo uma precisão de 98% para o melhor modelo e nunca confundindo *Series Temporais* de ruído com sinais (taxa de falsos positivos = 0). Notamos ainda que a maioria dos sinais classificados como ruído tinham razões sinal-ruído (SNRs) inferiores a 10. Além disso, uma rede adicional para regressão é implementada para estimar alguns parâmetros físicos do sistema, especificamente, a amplitude da GW (Δh) e a frequência máxima (pico) do sinal (f_{peak}), obtida durante o ressalto do núcleo, usando um conjunto de dados semelhante. Verifica-se que nos testes de regressão se obtêm melhores resultados nas previsões da f_{peak} do que para o Δh. A performance do treino está correlacionada com o SNR mínimo definido para o conjunto de dados: o melhor resultado foi obtido para um valor mínimo de SNR de 20, que usando Analise de Componentes Principais (PCA) nos gráficos de calibração dos valores previstos em função dos valores reais, fornece uma distribuição com variação $\sigma_{\Delta h} = 0.232$ e desvio do declive de 7.70% para o Δh e uma distribuição com variação $\sigma_{f_{\text{peak}}} = 0.174$ Hz e desvio do declive de 4.90% para a f_{peak}.

Palavras-chave: Classificação, Inferência de Parametros, Ondas Gravitacionais, Redes Neuronais, Supernovas de Colapso Nuclear

Abstract

Deep-Learning Inference of Rotational Core-Collapse Supernovae with Numerically-Generated Gravitational-Wave Signals

The gravitational collapse of the core of massive stars and the subsequent explosion of such stars as a Supernova (SN) is one of the most interesting sources of Gravitational Waves (GWs) expected to be detected in the coming years. Core-Collapse Supernovae (CCSNe) hold great scientific interest as the analysis of their complex waveforms can potentially provide valuable information about the underlying physical processes operating during the gravitational collapse of the iron cores of massive stars. In this work, Deep Learning (DL) techniques are used to analyze rotational CCSNe GW signals by classifying them and correctly separating them from the background noise of the detectors and by estimating some physical parameters of the systems that produce these signals.

To detect the presence of astrophysical GW from rotational CCSNe in noisy detector data, a classification network is trained using a dataset of *Time Series* from numerically generated CCSNe waveforms injected into the real noise of the Advanced LIGO and Advanced Virgo detectors (corresponding to the O3a observing run). The classification model achieves significant results, reaching an accuracy of 98% for the best model and never confusing strains of noise with signals (false positive rate = 0). It can also be noted that most of the injected signals classified as noise had Signal-to-Noise Ratios (SNRs) below 10. Moreover, an additional regression network is implemented to perform estimation on some physical parameters of the system, specifically, the GW strain amplitude (Δh) and the maximum (peak) frequency of the signal (f_{peak}), attained at core bounce, using a similar dataset. We find that the regression tests yield better results for the estimation of the f_{peak} than for Δh. The performance of the training was correlated with the minimum SNR set for the dataset: the best results were obtained for a minimum SNR of 20, which by using Principal Component Analysis (PCA) of the calibration plots of the predicted values vs real values gives a distribution's variation of $\sigma_{\Delta h} = 0.232$ and a slope deviation of 7.70% for Δh and a distribution's variation of $\sigma_{f_{\text{peak}}} = 0.174$ Hz and slope deviation of 4.90% for the f_{peak}.

Keywords: Classification, Core-Collapse Supernovae, Gravitational Waves, Neural Networks, Parameter Inference

Contents

List of Figures ix

List of Tables xii

Acronyms xiii

1 Introduction 1

2 Gravitational-Waves Signals 4

2.1 Gravitational Waves . . . 4

2.1.1 The Einstein-Hilbert Equations . . . 5

2.1.2 Linearized Theory . . . 7

2.2 Core-Collapse Supernovae Signals . . . 10

2.2.1 Supernovae Types and Mechanisms . . . 10

2.2.2 Equations Of State . . . 12

2.2.3 Supernovae Waveforms . . . 13

3 Gravitational-Waves Detection 16

3.1 The First Gravitational-Waves Detectors . . . 16

3.2 Current Detectors . . . 17

3.3 Future Detectors . . . 21

4 Deep Learning 23

4.1 Neural Networks . . . 24

4.1.1 Convolutional-Neural Networks . . . 26

4.1.2 Residual Blocks . . . 27

4.2 Regularization . . . 28

4.2.1 Cyclical Learning Rates . . . 30

5 Datasets and Methods **33**
5.1 Dataset Generation 33
5.1.1 Core-Collapse Supernovae Waveforms 33
5.1.2 Generating Time-Series Data 36
5.2 Neural Networks 38

6 Results and Discussion **41**
6.1 Classification Results 41
6.1.1 10k Dataset 41
6.1.2 100k Dataset 43
6.1.3 TimeSeries vs Spectrograms 44
6.2 Regression Results 46
6.2.1 10k Dataset 46
6.2.2 100k Dataset 48

7 Conclusions **56**

Bibliography **59**

List of Figures

1 Ring of particles deformed by the passage of a GW traveling in the direction of $\vec{K}$. The white dots and dashed lines represent the initial position of the particles in the ring, and the dark dots represent the effect of the GW. Only the "+" polarization is shown. The image was adapted from [20]. 5

2 The average blue-light curve for an observation set of SNII-P (a) and SNII-L (b), Figure 1 and 2 from [28]. 11

3 Scheme of Weber's Antenna. 16

4 Scheme of the interferometry method used in GWs detection. 18

5 Design sensitivity curve of Advanced LIGO (aLIGO) (a) and Advanced Virgo (AdV) (b) as presented in the official documents [80, 81]. 20

6 Design sensitivity curve expected for the two phases of AdV Plus (AdV+) in comparison with the sensitivity curve during third observation run (O3). 20

7 GWs sensitivity curves for the current detectors and future detectors and possible GWs sources to be detected. The image was produced with *http://gwplotter.com*. 22

8 Representation of a biological neuron. Image from [89]. 24

9 Representation of the *Cross-Correlation* operation in a convolutional layer. 27

10 Representation of a *Residual Block* by applying identity mapping. 28

11 Representation of three different models with the same training set. (Left) A linear function was fit to the data leading to *underfitting*. (Center) A quadratic function was fit to the data with no significant *overfitting* or *underfitting*. (Right) A polynomial of degree 9 function was fit to the data leading to *overfitting*. 29

12 Results of the classification of the dataset CIFAR-10, for different types of training: Original Learning Rate (LR), Exponential, and Cyclical Learning Rates (CLR). The plot shows the evolution of the accuracy of the different methods. These tests were performed by Leslie Smith in [94]. 31

13 Example of the *LR range test*, performed by Leslie Smith in [94]. These results were obtained for a *classification* test with an *AlexNet* network. 31

14 Results obtained by Smith and Topin in [95] by training a *Resnet-50* with the standard LR policy and with the *1cycle* policy. 32

15 Values of normalized f_{peak} as a function of normalized *maximum rotation rate* Ω_{max}. The dashed line represents the relation $2\pi f_{\text{peak}}/\sqrt{G\overline{\rho}_c} = 0.5(1 + \Omega_{\text{max}}/\sqrt{G\overline{\rho}_c})$. This figure was taken from [37] (Fig. 9). 34

16 Values of Δh as a function of $T/|W|$, showing a linear relation for $T/|W| \leq 0.06$. This figure was taken from [37] (Fig. 6). 35

17 18 waveforms for different Equation of State (EOS). The waveforms are plotted in the time domain for fixed values of the rotational profile parameters, $A = 634$km and $\Omega = 5.0$rads^{-1}. Taken from Richers *et al.* [37]. 36

18 Example of a dataset's element with an injected signal in the background noise from O3. 38

19 Example of a dataset's element of only background noise from O3. 38

20 Representation of the Residual CNN (ResCNN)'s architecture. 39

21 Evolution of the values of *Valid Loss* (a), *Accuracy* (b), and percentage of *False Negatives* (c) obtained when performing different *classification* tests with a varying number of *epochs*. 42

22 Confusion Matrix for the *classification* test with 25 *epochs* with a dataset of 10k *Time Series*, giving the number of strains predicted as noise or as signal versus their actual labels. . . 43

23 Receiver Operating Characteristic (ROC) Curve of the *classification* test with 25 *epochs* with a dataset of 10k *Time Series*. 43

24 Values of *Distance*, Δh, and f_{peak} as a function of the SNR for the missclassified strains on the dataset of 10k *Time Series*. The color bar represents the score given by the model to each strain. 44

25 Confusion Matrix for the *classification* test with 10 *epochs* with a dataset of 100k *Time Series*, giving the number of strains predicted as noise or signal versus their actual labels. . . . 44

26 ROC Curve of the *classification* test with 10 *epochs* with a dataset of 100k *Time Series*. . 45

27 Values of *Distance*, Δh, and f_{peak} as a function of the SNR for the missclassified strains for the dataset of 100k *Time Series*. The color bar represents the score given by the model to each strain. 45

28 *Confusion Matrix* from the work developed by Gabriel Mas with a dataset of 10k *spectrograms* [121]. 46

29 Calibration plots with the predicted Δh vs real Δh for values of $SNR \geq 5$ (a) and $SNR \geq 25$ (b), for a dataset of 10k *Time Series*. The white dashed line represents the points where the predicted value is equal to the real value $(y = x)$. 47

30 Calibration plots with the predicted f_{peak} vs real f_{peak} for values of $SNR \geq 5$ (a) and $SNR \geq 25$ (b), for the dataset of 10k *Time Series*. The white dashed line represents the points where the predicted value is equal to the real value $(y = x)$. 47
31 Probability distribution of the deviations of the Δh and f_{peak}, i.e., the difference between real value and predicted value, for the dataset of 10k *Time Series*. 48
32 Calibration plots of predicted Δh vs real Δh for values of $SNR \geq 5$ (a) and $SNR \geq 15$ (b) for a 100k dataset. 49
33 Calibration plots of predicted f_{peak} vs real f_{peak} for values of $SNR \geq 5$ (a) and $SNR \geq 15$ (b) for a 100k dataset. 49
34 Probability distribution of the deviations of the Δh for $SNR \geq 5$ (a) and $SNR \geq 15$ (b) for a 100k dataset. 50
35 Probability distribution of the deviations of the f_{peak} for $SNR \geq 5$ (a) and $SNR \geq 15$ (b) for a 100k dataset. 50
36 Values of *Valid Loss* (a), *Mean Absolute Error* (b) and *Root Mean Squared Error* (c) obtained when performing different *regression* tests with a different number of *epochs* using a dataset of 100k *Time Series* and setting the minimum SNR as 10. 51
37 Calibration plots of predicted Δh vs real Δh with $SNR \geq 5$ (a), $SNR \geq 10$ (b), $SNR \geq 15$ (c) and $SNR \geq 20$ (d) for the datasets of 100k *Time Series*. 52
38 Calibration plots of predicted f_{peak} vs real f_{peak} with $SNR \geq 5$ (a), $SNR \geq 10$ (b), $SNR \geq 15$ (c) and $SNR \geq 20$ (d) for the datasets of 100k *Time Series*. 53
39 Values of *Valid Loss* (a), *Mean Absolute Error* (b), *Root Mean Squared Error* (c), *Distribution Slope* (d), and *Distribution Variation* (σ) (e) obtained for the different regression tests for different values of the minimum SNR permitted in the datasets of 100k *Time Series*. . . 54
40 Probability distribution of the deviations of the Δh and f_{peak}, i.e., the difference between real value and predicted value for the dataset of $100k$ with a minimum SNR of 20 with datasets of 100k *Time Series*. 55

List of Tables

1 Range of parameters used to generate the datasets for classification and regression. . . 41

Acronyms

A+ aLIGO Plus *(pp. 19, 21)*
AdV Advanced Virgo *(pp. ix, 3, 18–20, 41)*
AdV+ AdV Plus *(pp. ix, 19, 20)*
AI Artificial Intelligence *(p. 23)*
aLIGO Advanced LIGO *(pp. ix, 3, 18–21, 41)*
ASD Amplitude Spectral Density *(p. 37)*
AUC Area Under the Curve *(pp. 40, 42, 44)*

BBH Binary Black Hole *(pp. 3–5, 12)*
BH Black Hole *(pp. 11, 12, 21)*
BNS Binary Neutron Star *(pp. 4, 5)*

CCSN Core-Collapse Supernova *(pp. vi, 3–5, 10–12, 15, 21, 33–35, 37, 38, 41, 42, 44–46, 49, 56–58)*
CE Cosmic Explorer *(pp. 21, 58)*
CFC Conformal Flatness Condition *(pp. 15, 35)*
CLR Cyclical Learning Rates *(pp. ix, 30, 31)*
CNN Convolutional Neural Network *(pp. 23, 25, 26, 38)*
cWB coherent Wave Burst *(pp. 49, 51, 56, 57)*

DL Deep Learning *(pp. vi, 3, 23–25, 30, 33, 38, 41, 44, 45, 56–58)*

EOS Equation of State *(pp. x, 10, 12, 13, 33–36, 56)*
ET Eintein Telescope *(pp. 21, 58)*

FCNN Fully Connected Neural Network *(pp. 24, 26)*

GR General Relativity *(pp. 4, 8, 10, 34, 35)*

GW Gravitational Wave *(pp. vi, ix, 2–5, 8–11, 16–18, 21, 22, 25, 33–36, 38, 41, 46, 50, 56, 57)*

KAGRA Kamioka Gravitational Wave Detector *(pp. 3, 17)*

LIGO Laser Interferometer Gravitational-Wave Observatory *(pp. 3, 17, 37)*

LISA Laser Interferometer Space Antenna *(p. 21)*

LR Learning Rate *(pp. ix, x, 25, 30–32, 39, 41, 43, 46, 51)*

LVK LIGO, Virgo and KAGRA *(pp. 21, 49)*

ML Machine Learning *(pp. 23, 28, 56)*

MLP Multilayer Perceptron *(p. 24)*

NN Neural Network *(pp. 3, 23–26, 29, 30, 56)*

NS Neutron Star *(pp. 3, 4, 11)*

O2 second observation run *(pp. 45, 56, 57)*

O3 third observation run *(pp. ix, x, 19, 20, 37, 38, 41, 50)*

O4 fourth observation run *(pp. 19, 20, 50, 51, 57)*

O5 fifth observation run *(pp. 19, 50, 51, 57)*

PCA Principal Component Analysis *(pp. vi, 40, 52, 57)*

PISN Pair-Instability Supernova *(p. 12)*

QRPN Quantum Radiation Pressure Noise *(pp. 18, 19)*

ResCNN Residual CNN *(pp. x, 23, 33, 38, 39, 56, 58)*

ResNet Residual Network *(pp. 25, 57)*

ROC Receiver Operating Characteristic *(pp. x, 40, 42–45)*

SN Supernova *(pp. vi, 2–5, 10, 11, 13, 21)*

SNR Signal-to-Noise Ratio *(pp. vi, x, xi, 42, 44–52, 54–57)*

WD White Dwarfs *(pp. 4, 10)*

1

Introduction

" (...) Behold, directly overhead, a certain strange star was suddenly seen, flashing its light with a radiant gleam and it struck my eyes. Amazed, and as if astonished and stupefied, I stood still, gazing for a certain length of time with my eyes fixed intently on it and noticing the same star placed close to the stars which antiquity attributed to Cassiopeia. When I had satisfied myself that no star of that kind had ever shone forth before. I was led into such perplexity by the unbelievability of the thing that I began to doubt the faith of my own eyes (...) " - Tycho Brahe, Progymnasmata, Chapter 3.

In 1006, a *new star* appeared in the night sky. It would stay visible for about three years and records about it were written all over the world. Its brightness was so intense that this *new star* could even be seen in the daytime [2]. In ancient times, when a bright star would suddenly appear in the sky, for a limited time, at a place where before was nothing, the star would be called a *guest star*. Transient objects were recorded extensively in China, Japan, and Korea, as it was customary to constantly observe the sky in search of signs about future occurrences. These were times when astronomy and astrology were highly intertwined.

The term *Nova* emerged as an abbreviation for the Latin expression *"Nova Stella"*, which means *new star*. The adoption of the name was spread by the works of Tycho Brahe on the 1572 *Nova*. This event was seen in Europe, Korea, and China. At the time, it was believed in Europe that the heavens were composed of eight celestial spheres, an idea dating back to the times of Aristotle. Tycho's measurements concluded that the *new star* had a fixed position, belonging to the eighth sphere with the "fixed" stars, the most distant of the spheres. This fact was not expected as they thought at the time that changes were only possible in the inner spheres, for regions closer than the Moon. On October 1604 another *Nova* appeared in the sky, usually associated with Kepler, as he made detailed observations, calculating its position to nearby stars very accurately. There are also records of this observation from China and Korea. These two events were studied in Europe by very influential astronomers with a special interest in understanding the intrinsic properties of the stars instead of only observing their position and movement, as was more common at the time. We could say that the studies of the events of 1572 and 1604 would lead to the

start of Astrophysics.

The term Supernova (SN) was first used by the Swedish Astronomer Knut Lundmark in his studies about the luminosities of *Novae*. Walter Baade and Fritz Zwicky followed his work in the early 1930s and defined a class of *Novae* with luminosities 10.000 times more intense than the others as SNe. *Novae* are usually associated with explosions occurring on the surface of stars. On the other hand, SNe are a consequence of the gravitational collapse of the progenitor star.

The study of gravitation and its laws had a major impulse in the 17th century when Sir Isaac Newton developed his law of universal gravitation. This law states that every massive object in the universe attracts every other massive object with a force that is directly proportional to the product of their masses and inversely proportional to the square of the distance between them. Newton's law was revolutionary at the time and allowed for the explanation of many astronomical phenomena, including the orbits of the planets around the sun.

In the 19th century, James Maxwell developed his mathematical field theory of electromagnetism and suggested that gravity could be treated in a similar way but ended up not being able to accomplish this goal. Following Maxwell's idea, Oliver Heaviside [3], in 1893, proposed a gravitational vector theory to describe the flux of gravitational energy for the first time. In 1905, Henri Poincaré wrote his formulation of relativity [4], proposing that gravity was carried by a wave. On that same year, Albert Einstein published his *Special Relativity Theory* [5], but the full version, the *Theory of General Relativity* [6], was only published in 1915, as a generalization of *Special Relativity*. In this version of his work, Einstein developed the mathematical formulation that showed how, contrary to Newton's theory, gravity was not a force but a geometrical effect of the curvature of spacetime caused by the presence of mass. Finally, after the insights Heaviside and Poincaré had given about gravity, Einstein's *Theory of General Relativity* provided field equations that, when linearized, could be cast in the form of a wave equation describing the propagation of a tensor perturbation of the flat Minkowski metric, i.e. Gravitational Wave (GW). Einstein's full theory was able to describe some observations that couldn't be explained by previous theories. One of the effects observed that was described for the first time by this theory was the perihelion motion of mercury [7]. In addition, Einstein's new gravitational theory could correctly measure the gravitational deflection of light by massive bodies, as famously observed for the first time during a solar eclipse in 1919 [8].

The first proof of the existence of GWs was obtained indirectly, in 1974. At that time, James Weber had already attempted to detect GWs directly using a resonant bar detector. Despite the initial claims, it is widely believed that Weber did not detect GWs as his instrument was not sensitive enough to observe them. The discovery of a *Binary Pulsar* by Taylor and Hulse, revealed that gravitational energy could be radiated by an accelerated mass [9]. Some years later, in 1979, the effects of GWs were measured in the *Binary Pulsar PSR 1913 + 16* by Taylor, Fowler, and McCulloch [10].

These discoveries propelled the development of new detectors and the construction of observatories for the direct detection of GWs. The idea of using a Michelson interferometer for GW detection was first explicitly suggested in the former USSR by Gertsenshtein and Pustovoit in 1962 [11], but this idea never came to fruition. Some years later, in 1971, Forward published the design of the first interferometer

prototype, following a suggestion of James Weber (1964) [12]. Independently, Weiss also projected a Laser Interferometer, inspired by the works of Pirani [13].

The Laser Interferometer Gravitational-Wave Observatory (LIGO) project [14] started construction in 1994 and only a few years later, in 1996, the Virgo [15] facilities also began their construction. The operations of *initial* LIGO and Virgo started several years later, in 2002 and 2007, respectively. After major upgrades of the instruments to their current *advanced* designs, on the 14th of September 2015 the first successful direct detection of GWs was accomplished by the two Advanced LIGO (aLIGO) detectors. This event, GW150914, was the product of the merger of a Binary Black Hole (BBH) system [16]. In 2017, the first multi-messenger detection was made when GWs from the merger of two Neutron Stars (NSs) were detected by the LIGO and Virgo detectors, followed by electromagnetic observations by gamma-ray, X-ray, optical, infrared, and radio telescopes. With each upgrade, these detectors can provide increasingly more precise measurements of GWs, allowing us to gain a greater insight into the astrophysical sources that generate them and into the Universe. During the first three observing runs, the number of confirmed detections of signals from compact binary coalescences has reached 90 events [17]. With the future upgrade plans for both aLIGO and Advanced Virgo (AdV), and with the addition of Kamioka Gravitational Wave Detector (KAGRA) [18] in the detector network, the number of detection is expected to rise. In the most recent optically-targeted searches for GW from Core-Collapse Supernovae (CCSNe) using data from the first and second observation runs of aLIGO and AdV no significant candidates were found [19]. As the sensitivity of the detectors increases the probability to detect the first GW signal from a CCSN also becomes higher. This event is much anticipated by the community, as the analyses of CCSNe waveforms could provide valuable information about the physical processes occurring during the collapse of massive stars.

This dissertation aims to investigate the use of Deep Learning (DL) methods to analyze GWs signals from rotational CCSNe. To gain a comprehensive understanding of the topic, Chapter 2 offers an introduction to GWs and their representation, as well as an overview of the various types of SNe and their mechanisms. Chapter 3 delves into GW detection, including a brief description of first, current, and future detectors. An introduction to DL will be given in Chapter 4, where the Neural Networks (NNs) will be presented. Chapter 5 details the methods tested in this work, and the results are presented and discussed in Chapter 6. Lastly, Chapter 7 will contain the final conclusions of the results and future directions of research.

2

Gravitational-Waves Signals

In this chapter, our focus will be on the theoretical studies of GWs. We will introduce different formalisms of General Relativity (GR) for the description of GWs and describe the different types of SNe and their mechanisms, as CCSNe are the sources of GWs analyzed in this dissertation.

2.1 Gravitational Waves

In 1915, Albert Einstein brought to light the *General Theory of Relativity* [6], describing how spacetime can be distorted by mass and how it influences the motions of objects. An underlying prediction of this theory is gravitational radiation. GWs are ripples in the fabric of spacetime traveling at the speed of light.

When a GW propagates through matter, the particles present will consequently move, undergoing accelerations on the plane perpendicular to the direction the wave propagates. For a better understanding of the effects of GWs on matter, we may consider a ring of particles at rest. When a wave passes perpendicularly to the plane of the ring, it will lose its round shape, elongating along one axis of the plane and contracting along the other, to then reverse this movement. This oscillatory motion, which has a frequency that coincides with that of the GW, can be visualized in Figure 1.

GWs can be originated from different types of phenomena, producing waves with distinct shapes and traits [21, 22]. *Stochastic GWs* have low amplitudes and are characterized by their random fluctuations. Such backgrounds are a consequence of the superposition of the GWs produced by a large number of unresolved, independent sources, much too weak to be individually detected with current technology [23]. *Stochastic GW* backgrounds are expected to have been created during the Big Bang or by the coalescence of BBHs and Binary Neutron Stars (BNSs) through cosmic history, even though other backgrounds of astrophysical origin exist for other types of sources, both transient, as SNe or mergers of White Dwarfs (WD) binaries, and continuous, as spinning NSs or magnetars. GWs that sustain their frequency and amplitude for a long time (almost continuous waves) are called *Periodic GWs*. They are originated by spinning NSs with a certain departure from axisymmetry in the distribution of their mass-quadrupole tensor and could also be generated by the accretion of matter by a NS in an X-ray binary system, triggering particular families of GW-emitting, oscillation modes in those stars. Very energetic events, such as the merger of

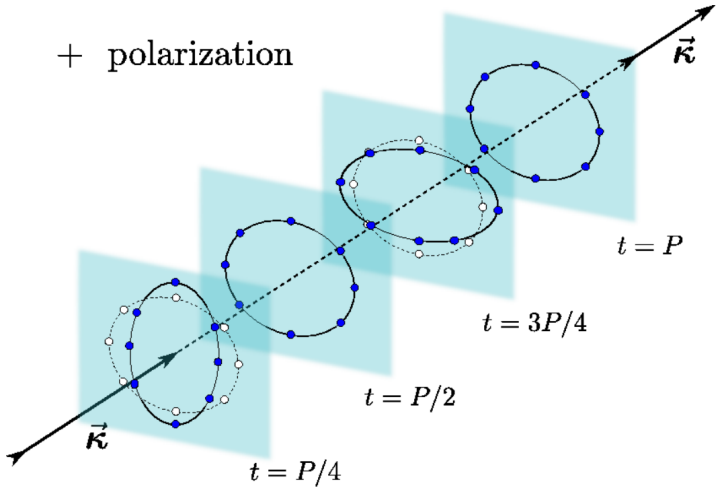


Figure 1: Ring of particles deformed by the passage of a GW traveling in the direction of $\vec{K}$. The white dots and dashed lines represent the initial position of the particles in the ring, and the dark dots represent the effect of the GW. Only the "+" polarization is shown. The image was adapted from [20].

BBHs or BNSs, generate chirp-like GWs, followed by a merger *burst-like GW* signal and a *ringdown* signal. *Burst GWs* are also associated with SNe explosions. These sources emit intense gravitational radiation in a very short period of time. The intensity of the GWs is higher for more massive sources, while their frequency is lower.

In order to better understand the physical processes that cause GWs, numerical models need to be developed. This involves solving the *Einstein-Hilbert Equations*, which are too complicated to solve in full generality. Only a few particular solutions that describe spacetimes with very restrictive conditions of symmetry have been calculated. For more specific cases, as is the case for CCSNe, we need to make some approximations. The use of *Numerical Relativity* is essential for these cases, as it provides new numerical solutions. We will see later in this chapter how *Numerical Relativity* can be used to solve the *Einstein-Hilbert Equations* for the specific case of CCSNe, but we will start by deducing and solving these equations for the most simple case first.

2.1.1 The Einstein-Hilbert Equations

The description of GWs can be taken by its linear approximation, where the *Einstein-Hilbert Equations* are linearised around the flat *Minkowski Metric*, $\eta_{\mu\nu}$. This section follows closely the discussions in [20, 24]. As a first step, we start by deriving the *Einstein-Hilbert Equations*. These equations can be obtained using the *Principle of Least Action* which takes the extremization of the action, requiring it to be stationary with respect to field variations, e.g. the variation of the action, S, must be equal to zero.

We will consider the *Gravitational Action* as the sum of the *Matter Action* (S_M) and the *Hilbert Action* (S_E), the term regarding the spacetime curvature,

$$S = S_M + S_E = \frac{1}{c}\int d^4x \left[\frac{c^4}{16\pi G}R + L_M\right]\sqrt{-g}, \tag{2.1}$$

where c and G are, respectively, the speed of light and the gravitational constant, g represents the determinant of the *Metric Tensor*, $g_{\mu\nu}$[1], and L_M represents the part of the Lagrangian describing the matter fields. In the second term of the equation, R represents the *Ricci Scalar*[2].

As we need to calculate the *Gravitational Action*'s variation, we can start by finding the *Matter Action* variation and *Hilbert Action* variation, separately. We can easily find the first one using the definition of the *Energy-momentum Tensor of Matter* $T_{\mu\nu}$, which is the variation of the *Matter Action* S_M under a change of metric $g_{\mu\nu} \longrightarrow g_{\mu\nu} + \delta g_{\mu\nu}$. This can be written as

$$\delta S_M = \frac{1}{2c} \int d^4x \sqrt{-g} T_{\mu\nu} \delta g_{\mu\nu}. \tag{2.2}$$

Now, we also need to calculate the *Hilbert Action*'s variation, which can be written as

$$\delta S_E = \frac{1}{c} \int d^4x \frac{c^4}{16\pi G} \delta(\sqrt{-g}R) = \frac{1}{c} \int d^4x \frac{c^4}{16\pi G} \left[R\delta(\sqrt{-g}) + \sqrt{-g}\delta R \right]. \tag{2.3}$$

To simplify the above equation we can calculate the variation of R, given by

$$\delta R = \delta(g^{\mu\nu} R_{\mu\nu}) = R_{\mu\nu}\delta g^{\mu\nu} + (g^{\mu\nu}\delta\Gamma^{\rho}_{\mu\nu} + g^{\mu\rho}\Gamma^{n}u_{\mu\nu})_{;\rho}, \tag{2.4}$$

and the variation of $\sqrt{-g}$, which can be written as

$$\delta\sqrt{-g} = -\frac{1}{2}\frac{\delta g}{\sqrt{-g}} = -\frac{1}{2}\frac{g g^{\mu\nu}\delta g_{\mu\nu}}{\sqrt{-g}} = -\frac{1}{2}\frac{(-g_{\mu\nu}\delta g^{\mu\nu})}{\sqrt{-g}} = -\frac{1}{2}\sqrt{-g} g_{\mu\nu}\delta g^{\mu\nu}. \tag{2.5}$$

Taking into consideration that δR is being integrated on a four-volume and its second term is a total derivative, it will not contribute to the dynamics. Using 2.4 and 2.5, we can write 2.3 as

$$\delta S_E = \frac{1}{c} \int d^4x \frac{c^4}{16\pi G} \left[(-R\frac{1}{2}\sqrt{-g} g_{\mu\nu}\delta g^{\mu\nu}) + (\sqrt{-g} R_{\mu\nu}\delta g^{\mu\nu}) \right]. \tag{2.6}$$

We can now combine 2.2 and 2.6 to write the variation of the *Gravitational Action* S as

$$\delta S = \delta S_E + \delta S_M = \tag{2.7}$$

$$= \frac{1}{c} \int d^4x \frac{c^4}{16\pi G} \left[(-R\frac{1}{2}\sqrt{-g} g_{\mu\nu}\delta g^{\mu\nu}) + (\sqrt{-g} R_{\mu\nu}\delta g^{\mu\nu}) \right] - \frac{1}{2c} \int d^4x \sqrt{-g} T_{\mu\nu}\delta g^{\mu\nu}, \tag{2.8}$$

and taking the *Principle of Least Action*, $\delta S = 0$, we have

$$\int d^4x \frac{c^4}{8\pi G} \left[(-\frac{1}{2}R\sqrt{-g} g_{\mu\nu}\delta g^{\mu\nu}) + (\sqrt{-g} R_{\mu\nu}\delta g^{\mu\nu}) \right] = \int d^4x \sqrt{-g} T_{\mu\nu}\delta g^{\mu\nu}, \tag{2.9}$$

which leads to the *Einstein-Hilbert Equations*:

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu}. \tag{2.10}$$

[1]The *Metric Tensor*, $g_{\mu\nu}$, is used in the form $(-+++)$.

[2]We define the *Riemann Tensor* as $R^{\mu}_{\nu\rho\sigma} = \partial_\rho\Gamma^{\mu}_{\nu\sigma} - \partial_\sigma\Gamma^{\mu}_{\nu\rho} + \Gamma^{\mu}_{\alpha\rho}\Gamma^{\alpha}_{\nu\sigma} - \Gamma^{\mu}_{\alpha\sigma}\Gamma^{\alpha}_{\nu\rho}$, where $\Gamma^{\rho}_{\mu\nu}$ represents the *Christoffel Symbols* as $\Gamma^{\rho}_{\mu\nu} = \frac{1}{2}g^{\rho\sigma}(g_{\sigma\nu;\mu} + g_{\sigma\mu;\nu} - g_{\mu\nu;\sigma})$. The *Ricci Tensor* is defined by $R_{\mu\nu} = R^{\alpha}_{\mu\alpha\nu}$ and the *Ricci Scalar* can be writen as $R = g^{\mu\nu}R_{\mu\nu}$.

2.1.2 Linearized Theory

In this section, we will expand the *Einstein-Hilbert Equation* around $\eta_{\mu\nu}$ to find their linear approximation which is called the *Linearized Theory*. For this, we can start by writing our metric as a small perturbation of a flat spacetime and then expand the equations of motion to linear order,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad |h_{\mu\nu}| \ll 1\,. \tag{2.11}$$

To write the *Einstein-Hilbert Equations* in their linearised version we will need to rewrite $R_{\mu\nu}$ and R by applying equation 2.11 and considering its derivative $g_{\mu\nu,\alpha} = \eta_{\mu\nu,\alpha} + h_{\mu\nu,\alpha} = h_{\mu\nu,\alpha}$ [3]. As $h_{\mu\nu}$ and $h_{\mu\nu,\alpha}$ are both small, we will consider from now on that the multiplications $h_{\mu\nu}h_{\alpha\beta}$, $h_{\mu\nu,\rho}h_{\alpha\beta}$ and $h_{\mu\nu,\rho}h_{\alpha\beta,\sigma}$ can be neglected (i.e. we only keep linear terms in the perturbations). With this in mind, we can write the *Christoffel Symbols* as $\Gamma^{\alpha}_{\mu\nu} = \frac{1}{2}\eta^{\alpha\sigma}(h_{\sigma\nu,\mu} + h_{\sigma\mu,\nu} - h_{\mu\nu,\sigma}) = \frac{1}{2}(h^{\alpha}_{\nu,\mu} + h^{\alpha}_{\mu,\nu} - h^{,\alpha}_{\mu\nu})$. This way, the *Ricci Tensor* at linear order can now be written as

$$R_{\mu\nu} = \Gamma^{\alpha}_{\mu\nu,\alpha} - \Gamma^{\alpha}_{\mu\alpha,\nu} = \frac{1}{2}\left((h^{\alpha}_{\nu,\mu\alpha} + h^{\alpha}_{\mu,\nu\alpha} - h^{,\alpha}_{\mu\nu,\alpha}) - (h^{\alpha}_{\alpha,\mu\nu} + h^{\alpha}_{\mu,\alpha\nu} - h^{,\alpha}_{\mu\alpha,\nu})\right)$$

$$= \frac{1}{2}\left(h^{\alpha}_{\nu,\mu\alpha} - h^{,\alpha}_{\mu\nu,\alpha} - h^{\alpha}_{\alpha,\mu\nu} + h^{,\alpha}_{\mu\alpha,\nu}\right).$$

Defining $h = h^{\mu}_{\mu} = \eta^{\mu\nu}h_{\mu\nu}$, and using the relation $\eta_{\mu\nu}\eta^{\mu\nu} = 1$, we can also write

$$R = \frac{1}{2}\left(h^{\alpha\mu}_{,\mu\alpha} - h^{\mu,\alpha}_{\mu,\alpha} - h^{\alpha,\mu}_{\alpha,\mu} + h^{,\alpha\mu}_{\mu\alpha}\right) = \frac{1}{2}\left(h^{\alpha\mu}_{,\mu\alpha} - h^{,\alpha}_{,\alpha} - \eta_{\alpha\mu}\eta^{\alpha\mu}h^{,\alpha}_{,\alpha} + \eta_{\alpha\mu}\eta^{\alpha\mu}\eta_{\mu\alpha}\eta^{\mu\alpha}h^{\alpha\mu}_{,\mu\alpha}\right) = h^{\alpha\mu}_{,\mu\alpha} - h^{,\alpha}_{,\alpha}.$$

We can now rewrite the *Einstein-Hilbert Equations*:

$$\left(h^{\alpha}_{\nu,\mu\alpha} - h^{,\alpha}_{\mu\nu,\alpha} - h^{\alpha}_{\alpha,\mu\nu} + h^{,\alpha}_{\mu\alpha,\nu}\right) - \eta_{\mu\nu}\left(h^{\alpha\beta}_{,\beta\alpha} - h^{,\alpha}_{,\alpha}\right) = \frac{16\pi G}{c^4}T_{\mu\nu}. \tag{2.12}$$

Although the equations are now linearized their mathematical structure still does not suggest the behavior of a wave. To accomplish this, we can first define the relation

$$\overline{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h \quad \Leftrightarrow \quad h_{\mu\nu} = \overline{h}_{\mu\nu} + \frac{1}{2}\eta_{\mu\nu}h. \tag{2.13}$$

Applying this to 2.12 we can simplify the expression above:

$$h^{,\alpha}_{\nu\alpha,\mu} + h^{,\alpha}_{\mu\alpha,\nu} - h^{,\alpha}_{\mu\nu,\alpha} - h_{,\mu\nu} - \eta_{\mu\nu}h^{,\alpha\beta}_{\alpha\beta} + \eta_{\mu\nu}h^{,\alpha}_{,\alpha} = \frac{16\pi G}{c^4}T_{\mu\nu}$$

$$(\overline{h}_{\nu\alpha} + \frac{1}{2}\eta_{\nu\alpha}h)^{,\alpha}_{,\mu} + (\overline{h}_{\mu\alpha} + \frac{1}{2}\eta_{\mu\alpha}h)^{,\alpha}_{,\nu} - (\overline{h}_{\mu\nu} + \frac{1}{2}\eta_{\mu\nu}h)^{,\alpha}_{,\alpha} - h_{,\mu\nu} - \eta_{\mu\nu}(\overline{h}_{\alpha\beta} + \frac{1}{2}\eta_{\alpha\beta}h)^{,\alpha\beta} + \eta_{\mu\nu}h^{,\alpha}_{,\alpha} = \frac{16\pi G}{c^4}T_{\mu\nu}$$

$$\overline{h}^{,\alpha}_{\nu\alpha,\mu} + \frac{1}{2}h_{,\mu\nu} + \overline{h}^{,\alpha}_{\mu\alpha,\nu} + \frac{1}{2}h_{,\mu\nu} - \overline{h}^{,\alpha}_{\mu\nu,\alpha} - \frac{1}{2}\eta_{\mu\nu}h^{,\alpha}_{,\alpha} - h_{,\mu\nu} - \eta_{\mu\nu}(\overline{h}^{,\alpha\beta}_{\alpha\beta} + \frac{1}{2}h^{,\alpha}_{,\alpha}) + \eta_{\mu\nu}h^{,\alpha}_{,\alpha} = \frac{16\pi G}{c^4}T_{\mu\nu}$$

$$\overline{h}^{,\alpha}_{\nu\alpha,\mu} + \overline{h}^{,\alpha}_{\mu\alpha,\nu} - \overline{h}^{,\alpha}_{\mu\nu,\alpha} - \eta_{\mu\nu}\overline{h}^{,\alpha\beta}_{\alpha\beta} = \frac{16\pi G}{c^4}T_{\mu\nu}$$

[3] As $\eta_{\mu\nu}$ is constant, its derivative will become zero.

$$\Box\overline{h}_{\mu\nu} + \eta_{\mu\nu}\overline{h}^{,\alpha\beta}_{\alpha\beta} - \overline{h}^{,\alpha}_{\mu\alpha,\nu} - \overline{h}^{,\alpha}_{\nu\alpha,\mu} = -\frac{16\pi G}{c^4}T_{\mu\nu}. \tag{2.14}$$

We can next use the *Lorenz gauge* to transform these equations further. Defining $\overline{h}^{,\nu}_{\mu\nu} = 0$, the linearised *Einstein-Hilbert Equations* take the form

$$\Box\overline{h}_{\mu\nu} = -\frac{16\pi G}{c^4}T_{\mu\nu}. \tag{2.15}$$

Equation 2.15 gives the basic result of the generation of GWs in *Linearized Theory*. This result takes into consideration a source of GWs that forms in flat spacetime. The use of the *Minkowski Metric*, $\eta_{\mu\nu}$, means that we are describing the dynamics using Newtonian gravity instead of considering full GR.

To look into the propagation of these GWs we can consider a region away from the source where $T_{\mu\nu} = 0$, which gives

$$\Box\overline{h}_{\mu\nu} = 0 \iff -(\frac{1}{c^2}\partial_0^2 + \nabla^2)\overline{h}_{\mu\nu} = 0. \tag{2.16}$$

Equation 2.16 suggests that GWs propagate at the speed of light, c.

If we consider a transformation of coordinates $x^\mu \longrightarrow x'^\mu = x^\mu + \xi^\mu(x)$, where the derivatives $|\xi_{\nu,\mu}|$ are of the same order as $|h_{\mu\nu}|$, at most, we can now use the transformation law of the metric,

$$g_{\mu\nu}(x) \longrightarrow g'_{\mu\nu}(x') = \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\sigma}{\partial x'^\nu}g_{\rho\sigma}(x), \tag{2.17}$$

to find the transformation of $h_{\mu\nu}$, to lowest order:

$$h_{\mu\nu}(x) \longrightarrow h'_{\mu\nu}(x') = h_{\mu\nu}(x) - (\xi_{\nu,\mu} + \xi_{\mu,\nu}). \tag{2.18}$$

Using 2.18, we can get the symmetry transformations:

$$\overline{h}_{\mu\nu} \longrightarrow \overline{h'}_{\mu\nu} = \overline{h}_{\mu\nu} - (\xi_{\nu,\mu} + \xi_{\mu,\nu} - \eta_{\mu\nu}\xi^\rho_{,\rho}) \quad and \quad \overline{h}_{\mu\nu,\nu} \longrightarrow (\overline{h}_{\mu\nu,\nu})' = \overline{h}_{\mu\nu,\nu} - \Box\xi_\mu, \tag{2.19}$$

and choosing a ξ_μ that leads to the transformation $\Box\xi_\mu = \overline{h}^{,\nu}_{\mu\nu}$. As we are considering the GWs outside of the source we can simplify the metric. If we keep $\Box\xi_\mu = 0$, we can use a coordinate transformation $x^\mu \longrightarrow x^\mu + \xi^\mu$ without altering the *Lorenz gauge*. As $\Box\xi_\mu = 0$ then $\Box\xi_{\mu\nu} = 0$, with $\xi_{\mu\nu} = \xi_{\nu,\mu} + \xi_{\mu,\nu} - \eta_{\mu\nu}\xi_{\rho,\rho}$.

Through the symmetry of the 4×4 matrix $h_{\mu\nu}$ we have ten independent components. As the *Lorenz gauge* gives four conditions, the number of independent components is reduced to six that will satisfy $\Box\overline{h}_{\mu\nu} = 0$.

Applying these coordinate transformations to each x^μ we can now choose the functions ξ^μ in a way that imposes another four conditions on $h_{\mu\nu}$. A good choice is a ξ^0 that leads to the trace $\overline{h} = 0$. If this happens, we also have $\overline{h}_{\mu\nu} = h_{\mu\nu}$. The other functions ξ^i are chosen so that $h^{0i}(x) = 0$. With these conditions the second equation in 2.19, with $\mu = 0$, will be:

$$\partial^0 h_{00} + \partial^i h_{0i} = 0. \tag{2.20}$$

If we fix $h_{0i} = 0$, we can simplify equation 2.20 to $\partial^0 h_{00} = 0$ so that h_{00} becomes constant in time. This indicates a static part of the gravitational interaction (the Newtonian potential of the source). The GW will be given by its time-dependent part. This will lead to the set of conditions called *Transverse-traceless gauge*, or *TT gauge*, defined in equations 2.21, i.e.,

$$h^{0\mu} = 0, \qquad h^i_i = 0, \qquad \partial^j h_{ij} = 0. \tag{2.21}$$

In the end, by imposing the *Lorenz gauge*, we reduced the ten degrees of freedom to six and then to only two. The metric in the *TT gauge* is denoted by h^{TT}_{ij}.

The equation $\Box\bar{h}_{\mu\nu} = 0$ admits a solution in the form of plane waves

$$h^{TT}_{ij}(x) = e_{ij}(\boldsymbol{k})e^{ikx}, \tag{2.22}$$

where $k^\mu = (\omega/c, \boldsymbol{k})$, with $\omega/c = |\boldsymbol{k}|$. $e_{ij}(\boldsymbol{k})$ is the *Polarization Tensor*. Since the condition $\partial^j h_{ij} = 0$ becomes $n^i h_{ij} = 0$, the non-zero components of h^{TT}_{ij} are in a plane transverse to $\hat{\boldsymbol{n}} = \boldsymbol{k}/|\boldsymbol{k}|$. If we chose $\hat{\boldsymbol{n}}$ along the z-axis, we obtain

$$h^{TT}_{ij}(t,z) = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}_{ij} cos[\omega(t - z/c)], \tag{2.23}$$

with h_{ij} symmetric and traceless. In a simpler way, we can also write

$$h^{TT}_{ab}(t,z) = \begin{pmatrix} h_+ & h_\times \\ h_\times & h_+ \end{pmatrix}_{ab} cos[\omega(t - z/c)], \tag{2.24}$$

where $a, b = 1, 2$ are the indices correspondent to the transverse plane (x, y). h_+ and $h_\times$ are the amplitudes of the *plus-polarization* and *cross-polarization* of the GW.

2.2 Core-Collapse Supernovae Signals

In this section, we will introduce CCSNe, the source of GWs we are interested in for this dissertation. We will start by describing the different types of SNe and the mechanisms that can lead to the explosion. Next, we will introduce some important theoretical concepts for the formulation of the CCSNe GWs: *Equation of State (EOS)*, *GR Hydrodynamics*, and the formulations of the metric equations.

2.2.1 Supernovae Types and Mechanisms

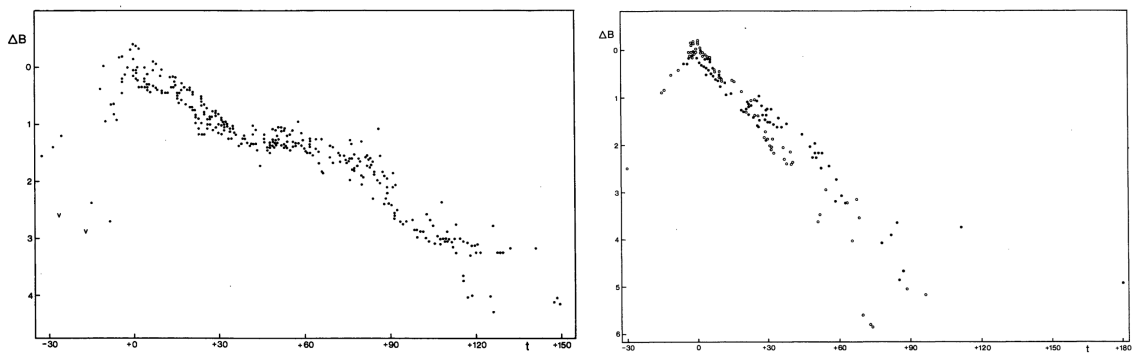
Some stars can develop into SN explosions at the end of their lives. The progenitor star can take different evolution paths, depending mainly on its initial mass, which leads to distinct SN mechanisms (see [25] for further details).

SNe are divided into two types through the analyses of their spectra [26]. If Hydrogen is not present in the spectrum, the SN is considered to be a *Type I*, called SNI. On the other hand, if there is Hydrogen, we classify it as *Type II SN*, called SNII. The absence of Hydrogen in some stars can be caused by nuclear fusion to heavier elements, by mass transfer in a binary system, or due to the presence of stellar winds that transport lighter elements into space. Looking further into the spectrum of SNI, we classify them as subtype *Type Ia SN* (SNIa) when Silicon is present. Besides the absence of Hydrogen, if Silicon is also not present, we distinguish the subtypes *Type Ib SN* (SNIb) and *Type Ic SN* (SNIc) with the presence or absence of Helium in the spectra, respectively [27]. There are also different subtypes for SNII. Considering the shape of the optical light curve of the event, a SN can be classified as a *Type II-P SN* (SNII-P) when, after 30-35 days of rapid declining, the luminosity reaches a *plateau*, and about 120-130 days later it decreases again at a reduced rate. When the luminosity's decrease is almost linear over time, the event is classified as a *Type II-L SN* (SNII-L) [28]. The shapes of these curves are shown in Figure 2. In addition, some events can be identified as *Type IIb SN* (SNIIb) when, in the beginning, their spectra are similar to SNII but, over time, they are more identical to SNIb or SNIc. When the SN spectrum is dominated by strong but narrow Balmer emission lines, the subtype is named *Type IIn SN* (SNIIn), where 'n' stands for "narrow" emission lines [29].

SNIa events are commonly associated with the explosion of WDs. A WD results from the evolution of stars with an initial mass between 0.5 to 8 $M_\odot$. They are rich in Carbon, Oxygen, and sometimes Silicon. Electron degeneracy pressure sustains the hydrostatic equilibrium in these stars. Such a configuration can only be possible for a limited range of mass. *Chandrasekhar* showed, in 1931, the mass limit to be:

$$M_{Ch} \approx \frac{5.83 M_\odot}{\mu_e^2} = 1.46 M_\odot \left(\frac{2}{\mu_e}\right)^2 \tag{2.25}$$

where μ_e is the mean molecular weight per electron [30, 31]. The *Chandrasekhar* limit may be surpassed when two WDs merge or, in a binary system, by the accretion of mass in one WD from its companion, leading to thermonuclear explosions. These events do not leave a stellar remnant.



(a) SNII-P (b) SNII-L

Figure 2: The average blue-light curve for an observation set of SNII-P (a) and SNII-L (b), Figure 1 and 2 from [28].

Another mechanism that can give rise to a SN explosion is the core collapse of a *Red Supergiant Star* with an initial mass between about 8 and 140 $M_{\odot}$. Most of these massive stars burn their thermonuclear fuel until they reach an Iron-group nucleus surrounded by lighter elements. The Iron core is stable due to the pressure of relativistic degenerate electrons. If the core exceeds its effective *Chandrasekhar Mass*, the collapse will start. Following the collapse, material is ejected through the progenitor's body. As progenitors can have different envelope structures, different progressions are expected of the SN's photosphere as material passes through, causing distinct light curves (SNII-P and SNII-L).

The collapse of the core takes less than 1 second. The temperature of the in-falling material rises and most of the collapse energy is released as neutrinos. This is the first sign that a CCSN event has occurred. The ejection of neutrinos contributes to the increase in temperature of the envelope material, which is discharged into space as SN ejecta. If the collapse is non-spherical, GWs are emitted from the core.

When a CCSN occurs, it can leave behind a remnant depending on its initial mass. If the progenitor's mass is below 40 $M_{\odot}$, a NS is formed. If the mass is higher than 40 $M_{\odot}$, it becomes a Black Hole (BH). The 40 $M_{\odot}$ limit can change depending on the progenitor's metallicity.

Shock waves are generated when the in-falling material bounces on the hard surface of a NS, taking about one hour to reach the outside of the star. The shock abruptly heats the progenitor's surface to a temperature of the order of one million degrees, emitting UV radiation. In only a day, the outer layers expand to 10 times the original size, and the brightness increases by a factor of 1000. At this point, the SN can be observed by simply looking into its electromagnetic spectrum. In the following weeks, the maximum brightness is reached. While its expansion continues and the outer layers diminish their density, the photosphere will eventually contract, leading the star's brightness to fade.

When a BH is formed, the in-falling material will not find a hard surface where it can bounce, so no shock waves travel out from the progenitor. As the energy in the outer layers is reduced, the SN will have a fainter brightness. There could even be a case where no remnant cloud is created, as all the material

from the progenitor star could be absorbed by the BH. In extreme cases, a star could suddenly disappear in the sky by the collapse of an entire star without any significant outflow of material.

For masses above 120 $M_\odot$, stars have an interstellar medium composition distinct from the Sun. In this case, the star's luminosity is higher than the *Eddington limit*, and its hydrostatic equilibrium is compromised as the radiation pressure overcomes gravitational attraction. For low metallicities, where the material is more transparent, this limit can be raised to 150 $M_\odot$ or more. In this case, we have a Pair-Instability Supernova (PISN). The name derives from the electron-positron pair-instability state created in the core of these stars during Oxygen burning, preventing a stable Iron core from forming. The thermal pressure decreases in its core due to pair production caused by nuclei and gamma rays' collisions. This mechanism does not leave a remnant behind [4].

2.2.2 Equations Of State

When matter is compressed during gravitational collapse to reach the point of nuclear saturation density or even goes beyond it, the behavior of matter is known with much uncertainty. Many different EOS have been proposed to attempt the characterization of this behavior [34, 35]. The EOS is fundamental to the dynamical evolution of CCSNe and the structure of the compact stars produced. For this reason, the EOS is an essential ingredient for the simulations of these events.

At temperature values near 0 K, the most relevant information is the relation between the pressure (P) and energy density. Considering matter in weak-interaction equilibrium, its composition is usually parameterized by the number of electrons per baryon (electron fraction Y_e). As an example, we show the pressure dependencies near nuclear saturation density n_s in equation 2.26, where n represents the nucleon number density and E is the binding energy per baryon,

$$P(n, Y_e) = n^2 \frac{\partial E(n, Y_e)}{\partial n}. \tag{2.26}$$

E can be written as a *Taylor* expansion for uniform nuclear matter. At zero temperature and around n_s of symmetric matter ($Y_e = 0.5$), E is defined as

$$E(x, \beta) = -E_0 + \frac{K}{18}x^2 + \frac{K'}{162}x^3 + ... + S(x, \beta), \tag{2.27}$$

where $E_0 \approx 16$ MeV (the bulk binding energy of symmetric matter at n_s), x is defined as a function of n as $x = (n - n_s)/n_s$, and $\beta = 1 - 2Y_e$. Moreover, K is the nuclear incompressibility modulus, K' is called the skewness parameter, and $S(x, \beta)$ is the symmetry term that contains the effects of Y_e. $S(x, \beta)$ is expanded around symmetric matter which only has even order elements as a consequence of the charge invariance of the nuclear interaction: $S(x, \beta) = S_2(x)\beta^2 + S_4(x)\beta^4 + ... \approx S_2(x)\beta^2$.

[4] The fact that PISNe do not leave a remnant leads to a gap in the mass distribution of BHs. The detection of GW190521 is consistent with a BBH merger where the primary BH as a mass that falls inside the PISN mass gap. This detection raised concerns about the actual nature and formation of the binary system that led to GW190521 [32] and also led into investigations of other factors that could change the boundaries of the mass gap [33].

We can define the saturation density, n_s, as the points where $dE(x,\beta)/dx = 0$. At densities above n_s, protons and electrons are uniformly distributed such that *Coulomb* effects are not important. The values for n_s, E_0, and mass density (ρ_s) are well constrained at the point of saturation density from laboratory measurements and theoretical calculations [36], which are consistent with estimates from astrophysical observations. These developments are essential as they decrease the range of EOS properties used in SNe simulations.

2.2.3 Supernovae Waveforms

The waveforms used in this dissertation were performed by [37] through numerical simulations using the `CoCoNut Code` [38]. In this subsection, we will present the formalism used in this algorithm, as described in [39].

2.2.3.1 General Relativistic Hydrodynamics

Considering a relativistic perfect fluid, with four-velocity u^μ and rest-mass density ρ, we can write its hydrodynamic evolution in a dynamic spacetime $g^{\mu\nu}$ using the following pair of local conservative equations:

$$\nabla_\mu J^\mu = 0 \qquad and \qquad \nabla_\mu T^{\mu\nu} = 0, \tag{2.28}$$

where $J^\mu = \rho u^\mu$ is the rest-mass current and $T^{\mu\nu} = \rho h u^\mu u^\nu + P g^{\mu\nu}$ is the energy-momentum tensor, with the specific enthalpy defined as $h = 1 + \epsilon + P/\rho$, being ϵ the specific internal energy. The three-velocity of the fluid, v^i, is defined as

$$v^i = \frac{u^i}{\alpha u^0} + \frac{\beta^i}{\alpha},$$

where α is the lapse function and β^i is the shift vector, two kinematic variables related with the 3+1 formulation of *Einstein-Hilbert Equations* (see e.g. [40] for details; see also below).

Following [41] let us define a set of conserved variables in terms of ρ, v_i, and ϵ

$$\begin{aligned} D &\equiv \rho W, \\ S_i &\equiv \rho h W^2 v_i, \\ \tau &\equiv \rho W^2 - P - D, \end{aligned} \tag{2.29}$$

where the term W is the *Lorentz factor* defined as $W = \alpha u^0$. Using these variables, we can write the *Flux-Conservative Hyperbolic System Equations* as

$$\frac{1}{\sqrt{-g}}\left[\frac{\partial\sqrt{\gamma}\mathbf{U}}{\partial t} + \frac{\partial\sqrt{-g}\mathbf{F}^i}{\partial x^i}\right] = \mathbf{Q}, \tag{2.30}$$

to represent the local conservation laws, with the *State Vector* **U**, *Flux Vector* **F**, and *Source Vector* **Q** respectively given by

$$\begin{aligned}
\mathbf{U} &= \left[D, S_j, \tau\right], \\
\mathbf{F}^i &= \left[D\hat{v}^i, S_j\hat{v}^i + \delta^i_j P, \tau\hat{v}^i + Pv^i\right], \\
\mathbf{Q} &= \left[0, T^{\mu\nu}\left(\frac{\partial g_{\nu j}}{\partial x^\mu} - \Gamma^\lambda_{\mu\nu} g_{\lambda j}\right), \alpha\left(T^{\mu 0}\frac{\partial ln\alpha}{\partial x^\mu} - T^{\mu\nu}\Gamma^0_{\mu\nu}\right)\right].
\end{aligned} \tag{2.31}$$

In these vectors, $\hat{v}^i = v^i - \beta^i/\alpha$, and $\sqrt{-g} = \alpha\sqrt{\gamma}$. This formalism, known as the *Valencia formulation of general-relativistic hydrodynamics*, was developed by Banyuls *et al.*, and full details can be found in [41].

2.2.3.2 The Metric Equations

For the metric, Dimmelmeier *et al.* used an adaptation of the *ADM 3 + 1 formalism* [40, 42], named after its authors R. Arnowitt, S. Deser, and C. W. Misner. This adaptation consisted in separating the spacetime into different non-intersecting spacelike hyper-surfaces. We can write the line element as:

$$ds^2 = -\alpha^2 dt^2 + \gamma_{ij}(dx^i + \beta^i dt)(dx^j + \beta^j dt), \tag{2.32}$$

where α describes the rate of time advance along a timelike unit vector, normal to the hyper-surface, and β^i describes the motion of coordinates within the surface. In this formalism, the *Einstein-Hilbert Equations* are split into four groups: the evolution equations for γ_{ij}, the evolution equations for the extrinsic curvature K_{ij}, the *Hamiltonian* constraint equations, and the momentum constraint equations. These equations have to be fulfilled at every spacelike hyper-surface and can be written as

$$\begin{aligned}
\partial_t\gamma_{ij} &= -2\alpha K_{ij} + \nabla_i\beta_j + \nabla_j\beta_i, \\
\partial_t K_{ij} &= -\nabla_i\nabla_j\alpha + \alpha\left(R_{ij} + KK_{ij} - 2K_{ik}K^k_j\right) + \beta^k\nabla_k K_{ij} + K_{ik}\nabla_j\beta^k + K_{jk}\nabla_i\beta^k \\
&\qquad - 8\pi\alpha\left(S_{ij} - \frac{\gamma_{ij}}{2}\left(S^k_k - \rho_H\right)\right), \\
0 &= R + K^2 - K_{ij}K^{ij} - 16\pi\rho_H, \\
0 &= \nabla_i\left(K^{ij} - \gamma^{ij}K\right) - 8\pi S^j.
\end{aligned} \tag{2.33}$$

The ADM equations have been shown to be numerically unstable due to being only weakly hyperbolic. Many studies have been performed to reformulate these equations into a strongly hyperbolic system, with the preferred current approach being the so-called BSSN system (see details in [40, 43]). Following a different approach, Dimmelmeier *et al.* concluded it was generally considered that the more constraints used in the formulation of the equations the more numerically stable the evolution would be. As done in their previous work [44], they decide to assume, during the numerical simulation, that all off-diagonal components of the three-metric are zero and that the diagonal elements have a common factor ϕ^4 that depends on the time and space coordinates. To accomplish this, they approximate the metric $g_{\mu\nu}$ by replacing the *spacial three-metric* γ_{ij} with the *conformally flat three-metric* $\gamma_{ij} = \phi^4\hat{\gamma}_{ij}$, where $\hat{\gamma}_{ij}$ is the

flat metric. This is called the Conformal Flatness Condition (CFC) approximation for the spatial metric. In the CFC approximation, we can write the relation $\partial_t \phi = \frac{\phi}{6} \nabla_k \beta^k$. With this, K_{ij} becomes time-independent and can be written as:

$$K_{ij} = \frac{1}{2\alpha} \left(\nabla_i \beta_j + \nabla_j \beta_i - \frac{2}{3} \gamma_{ij} \nabla_k \beta^k \right).$$

Employing the maximal slicing condition, $K = 0$, the ADM equations are reduced to a set of five coupled, elliptic, nonlinear equations for the metric components,

$$\begin{aligned} \hat{\Delta}\phi &= -2\pi\phi^5 \left(\rho h W^2 - P + \frac{K_{ij}K^{ij}}{16\pi} \right), \\ \hat{\Delta}(\alpha\phi) &= 2\pi\alpha\phi^5 \left(\rho h (3W^2 - 2) + 5P + \frac{7K_{ij}K^{ij}}{16\pi} \right), \\ \hat{\Delta}\beta^i &= 16\pi\alpha\phi^5 S^i + 2\phi^{10} K^{ij} \hat{\nabla}_j \left(\frac{\alpha}{\phi^6} \right) - \frac{1}{3} \hat{\nabla}^i \hat{\nabla}_k \beta^k, \end{aligned} \tag{2.34}$$

where $\hat{\nabla}_i$ and $\hat{\Delta}$ are, respectively, the flat space *Nabla* and *Laplace* operators. These equations have not explicit time derivatives so the metric is calculated in a fully constrained way. Therefore, the evolutionary degrees of freedom in the metric are neglected. For the specific case of CCSNe simulations, the equations are entirely dominated by the source hydrodynamic quantities (ρ, P, and v^i), and the nonlinear coupling only becomes important for strong gravity. Given this fact, on each time slice the metric is only determined by the instantaneous hydrodynamic states, i.e. the distribution of matter in space. The CFC system of equations has been extended [45] and other formulations have been used [43] but the results obtained are similar to those of Dimmelmeier *et al.* [44]. In [46], a direct comparison has also been done between the CFC approximation and perturbative analytical approaches (*post-Newtonian* and *effective-one-body*) and they are shown to have a robust agreement.

3

Gravitational-Waves Detection

This chapter outlines the evolution of GWs detection. We begin by examining the initial detectors developed for this purpose, followed by a summary of the techniques employed by the current detectors. Lastly, we will explore some of the detectors planned for the future.

3.1 The First Gravitational-Waves Detectors

The first GWs detector was developed by Joseph Weber. In 1966, he published the details of his "antenna" [47]. Weber's detector was composed of a large aluminum cylinder weighing three tons, suspended from a steel wire within a vacuum chamber to filter out external oscillations. The measurement of the vibrations induced on the detector was obtained with *piezoelectric crystals*, circling the body of the cylinder. In theory, the crystals would sense the motions induced by GWs and convert these mechanical vibrations into electrical impulses. The scheme of Weber's Antenna is represented in Figure 3.

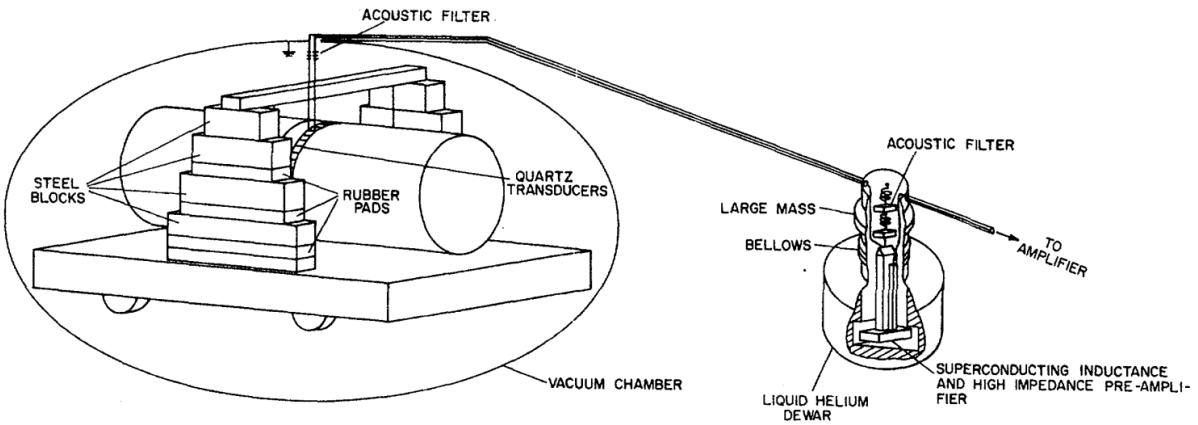


Figure 3: Scheme of Weber's Antenna.

Weber constructed two detectors and strategically placed them in two different laboratories, 950 km apart. The detection would only be considered valid if the same signal was detected simultaneously in both locations. This was a great advantage, allowing Weber to detect signals from distant sources in a

more accurate and reliable way, as it assured that the signals were not due to any local activity, such as seismic movements or thunderstorms, but instead were from an external source.

Some years later, the detector would be upgraded with isolation from local earthquakes and electromagnetic interference. Weber affirmed that the only significant source of noise in his cylinder, at that point, was from the thermal motions of its atoms, which produced fluctuations in the length of the cylinder around 10^{-16}m. Unfortunately, it was not likely that a GW signal would have a much greater amplitude than this thermal agitation, hence the detector was, most likely, not sensitive enough to detect GWs signals.

The results of the signals detected by Weber's antenna were only published in 1969 [48]. A year later, Weber claimed that several GW signals had been detected from the center of the *Milky Way* but these claims were met with much skepticism by the scientific community.

In the years that followed, Weber's design was replicated and improved upon by many other research groups. Some of these enhanced detectors were cooled to reduce the thermal noise [49]. Despite these improvements, it has never been possible to detect GWs with a technique based on resonant bars, and the detections claimed by Weber were considered spurious.

3.2 Current Detectors

Among other setups, the interferometer method was also proposed for the detection of GWs. The first detector to use this method was built in 1978 by a former student of Weber, Robert L. Forward [50]. Presently, several highly sophisticated GWs interferometers are working in different parts of the world. These detectors are capable of providing incredibly precise measurements of GWs. The two main detectors are the LIGO [14] and the *Virgo Interferometer* [15]. LIGO has two detectors in the United States, one in Hanford, Washington, and one in Livingston, Louisiana, both with two perpendicular arms of 4km length. Virgo is located in Italy and its arms are 3km long. Other detectors are available, such as *GEO600* [51], in Germany, and the KAGRA detector [18], in Japan. *GEO600* is a smaller detector with only 600 m arms. New technologies are developed and tested in *GEO600* before being implemented in the larger detectors. While LIGO, Virgo, and *GEO600* are detectors placed on the Earth's surface, KAGRA is an underground detector with 3km long arms.

These laser interferometer detectors aim to measure the difference in the length of two perpendicular arms caused by the passage of a GW. Its design is based on a *Michelson-Morley Interferometer* and its scheme is represented in Figure 4. The length of the detector's arms determines the type of signals that would be within the instrument's frequency sensitivity range and is chosen depending on the frequency of the GWs amenable to be detected. The optimal size would be one-fourth of the wavelength. For the case of GWs with 100Hz of frequency, it would imply a 750 km arm (L), which is not reasonable to construct. To overtake this problem, modern detectors use *Fabry-Pérot resonant cavities*, in which the laser beams go back and forth along the arms until they travel the desired length.

This method involves the emission of a monochromatic-laser beam which is divided equally into two

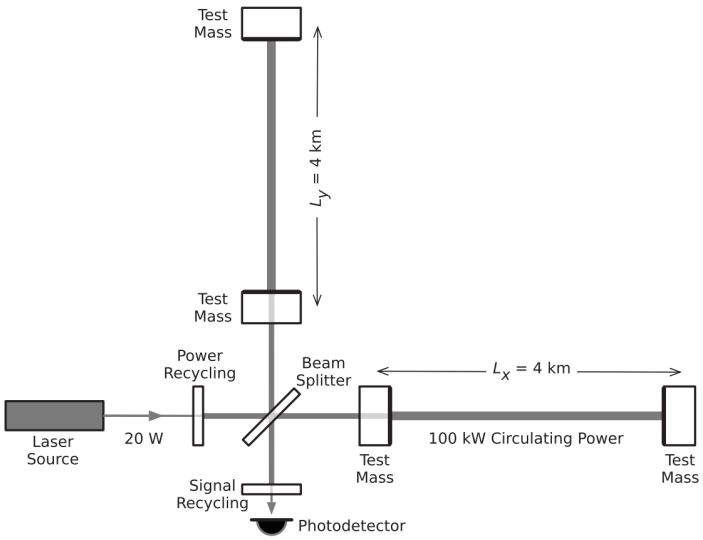


Figure 4: Scheme of the interferometry method used in GWs detection.

by a beam splitter. These beams travel in orthogonal directions along the two arms of the interferometer reflecting back and forth hundreds of times with the help of highly reflective mirrors. It is critical to have a very powerful beam of light traveling along the arms. To increase the effective laser power, a power recycling mirror [52] is installed between the laser source and the beam splitter.

As a GW passes through the detector, the length of the arms of the interferometer slightly changes. This change in length causes the beam of light traveling along one arm to travel a different distance from the beam of light traveling along the other arm. A quantity Δl amounts to the difference in distance between the beam splitter and one of the mirrors with respect to the distance between the beam splitter and the second mirror, $\Delta l = \delta L_x - \delta L_y$. This difference in the paths of the two beams causes interference in the light when they are recombined at the beam splitter (such an interference is perfectly destructive in the absence of a passing GW signal). The recombined beam is then reflected toward a photodetector where its intensity is measured. By measuring the amount of interference, we can determine how much the length of the arms changed, and thus detect the presence of a GW. We define the amplitude h of the GW as the maximum displacement per unit length. This dimensionless amplitude is called the strain and it is expressed by the relation $h = \Delta l / L$ and has an order of magnitude of $\sim 10^{-21}$.

Present detectors have improved substantially in the last few years with the upgrade from the initial instruments to their second-generation, advanced versions, aLIGO [53] and AdV [54]. To be able to detect GWs, it is imperative to reduce noise from external sources as much as possible. To achieve the required high sensitivities set by the typical strains of a GW, all sources of noise have to be taken into account and a collection of strategies have to be implemented to overcome them.

Quantum Noise is one of the sources that deeply limits the sensitivity of the detectors [55]. This phenomenon arises from fluctuations of the vacuum electric field during the measurements in the interferometer. It may appear as *shot noise* or Quantum Radiation Pressure Noise (QRPN). *Shot noise* derives from the statistical fluctuations of the arrival time of photons on the photodetector, and it appears

in the spectrum on higher frequencies. This effect can be reduced with the injection of squeezed vacuum states [56]. On the other hand, QRPN is generated by amplitude fluctuations of the electric field in the arms. The radiation pressure applied to the optics produces a momentum oscillation, inducing displacement noise. QRPN has higher importance at lower frequencies [57, 58].

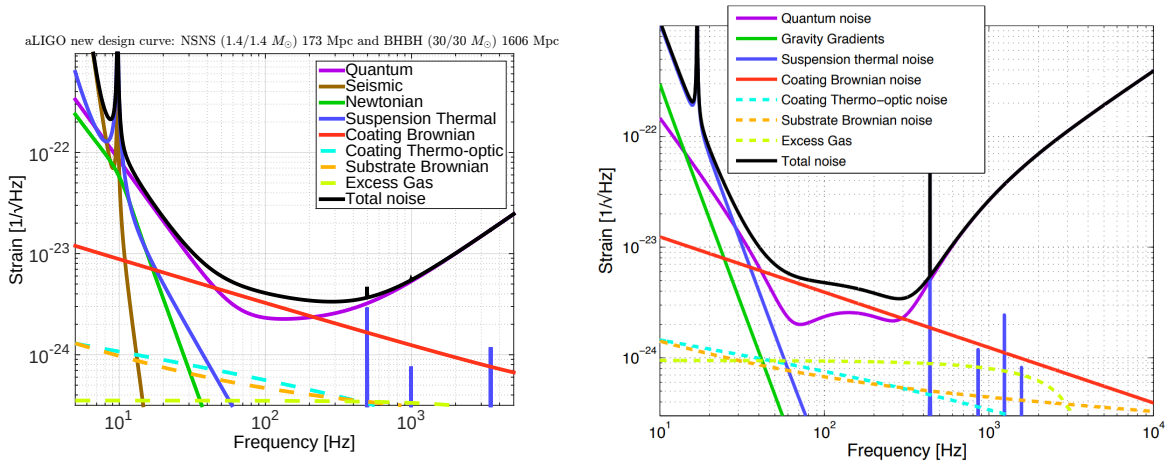
Another important source of noise, already recognized by Weber in his antenna, is *Thermal Noise*. The thermal motion of the atoms in the materials composing the test masses (mirrors) [59–61] and its suspension system causes displacement noise. To reduce the Brownian motions, low-loss optical coatings and cryogenic optics can be applied. The aLIGO detector uses Titania-doped Tantala or Silica coatings, varying the coating layers' thicknesses [62–64]. The application of fused silica substrate material on the suspension system also helps to lower the mechanical loss and has a small contribution to *thermal noise*.

The suspension of the test masses also dates back to Weber, with the goal of reducing *Seismic Noise*. These perturbations are caused by all forms of ground motion, wind, or human activity. The design of the aLIGO detectors involves a quadruple suspension system to reduce the influence of these ground motions on the test masses [65]. This system is supported by a hydraulic actuator structure [66] with seismic isolation platforms [67]. On the other hand, in Virgo is applied a different system for the same goal, the *Superattenuators* [68]. This system is composed of a five-stage pendulum suspending a last stage (the payload) that includes the test mass. On top, supporting the suspension system, is a platform ring connected to the ground by three legs with elastic joints to further reduce horizontal vibrations. With these implementations, the seismic noise stays far below the noise curve of the detectors.

Test masses can also suffer oscillations produced by terrestrial density fluctuations near the detectors. These effects are called *Newtonian Noise* or *Gravity Gradient* [69, 70] and are a consequence of seismic waves or atmospheric mass movements that produce changes in the local gravitational forces applied to the test masses [71]. It affects the sensitivity of the detectors at lower frequencies, under 20 Hz. These perturbations can be corrected by accurately measuring and modeling the seismic motions and later performing noise cancellation [72].

There are other sources of noise but their contributions are not as significant as the ones presented in the previous discussion. For a complete description of the sensitivity and performance of aLIGO and AdV during the last observing run of the two detectors, the third observation run (O3), the interested reader is addressed to [68, 73]. In Figure 5, we can see how each source of noise contributes to the design sensitivity curves of aLIGO and AdV.

Right after the conclusion of the O3 run, many improvements and upgrades started to be implemented to increase the sensitivity of the current detectors [68, 74]. These upgrades are the first phase of construction for the detectors' new versions, aLIGO Plus (A+) [75] and AdV Plus (AdV+) [76]. The LIGO-Virgo-KAGRA (LVK) Collaboration will start the fourth observation run (O4) observing run in May 2023 [77], which will have a duration of 18 months, but the full upgrade of the detectors to their respective design sensitivities will only be realized for the fifth observation run (O5). The major goal of these first upgrades is to increase the laser power at the source, install a *signal recycling cavity*, and implement *frequency-dependent squeezing* techniques [78]. With these implementations, the limit imposed by *Quantum Noise*



(a) aLIGO (b) AdV

Figure 5: Design sensitivity curve of aLIGO (a) and AdV (b) as presented in the official documents [80, 81].

will be reduced. After O4, the goal is to enlarge the size of the beam when meeting the mirrors, increase the test masses up to 105 kg and apply new highly reflective coatings [79], leading to a decrease in *Thermal Noise*. These improvements in the sensitivity of the detectors will permit the detection of fainter signals. The detection rate is expected to raise by a factor of 5. In Figure 6, we can see how the sensitivity for AdV+ is expected to increase.

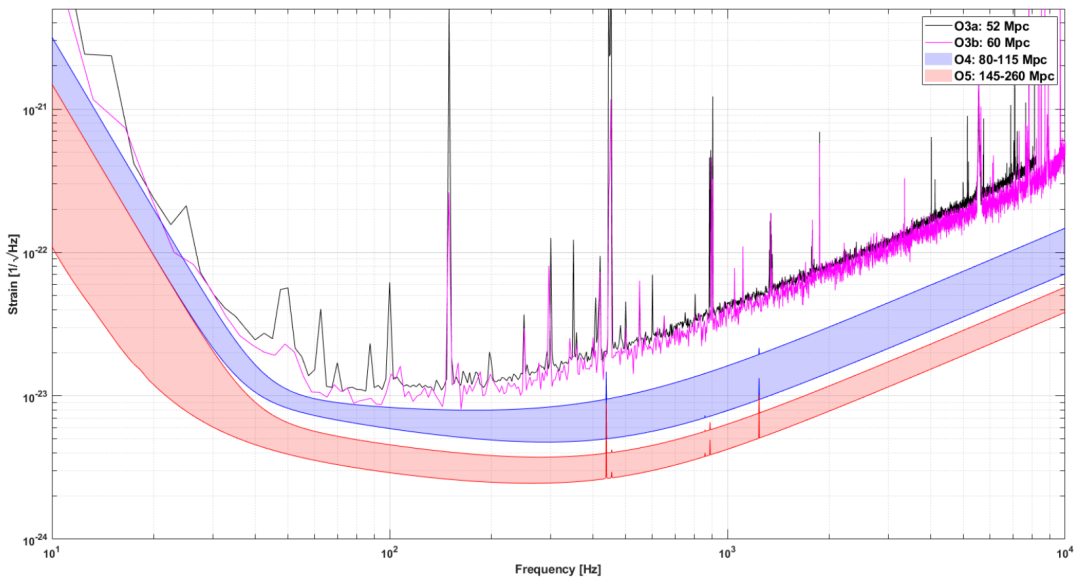


Figure 6: Design sensitivity curve expected for the two phases of AdV+ in comparison with the sensitivity curve during O3.

3.3 Future Detectors

In order to further advance the field of GWs detection, GWs detectors are continually being upgraded and new ones are being developed. A new detector similar to aLIGO/A+ is expected to come online in the future (within this decade), the *LIGO India Observatory* [82]. With another detector in the LIGO, Virgo and KAGRA (LVK) network, it will be easier to determine the location of the GWs sources.

The Laser Interferometer Space Antenna (LISA) project [83] is a space-based antenna that will offer a new window of observation into much lower frequencies than current ground-based detectors allow. It consists of three spacecrafts arranged in an equilateral-triangle formation, each carrying two cubes housing a laser interferometer. These cubes would detect very small changes in the distance between them caused by GWs passing through the detector. With LISA we expect to measure frequencies between 0.03 mHz and 0.1 Hz and detect GWs from the coalescence of very massive BHs at high redshifts, as well as waves from tens of stellar-mass compact objects spiraling into central massive BH. LISA is expected to be launched in 2037 [84].

The Eintein Telescope (ET) [85] is a proposed third-generation ground-based detector to be built in Europe. It would employ a triangle-shaped, underground, cryogenic interferometer with arms 10 km long, offering greater sensitivity than current second-generation detectors. It would measure a broader range of frequencies, from 0.1 Hz to 10 kHz, and detect GWs from compact binary inspirals, SNe, CCSNe, and gamma-ray bursts. It is expected that ET will begin observations around 2035 [86].

The Cosmic Explorer (CE) [87] is another proposed third-generation, ground-based, cryogenic, detector. It would feature arms 40 km long, enabling slightly greater sensitivity than the ET. In terms of frequencies, it is expected to observe GWs in the same window as ET.

Figure 7 shows the sensitivity curves expected for these future projects and the curves for current detectors. It allows us to compare their sensitivities and visualize how much they are expected to increase. The figure also shows the frequencies of different sources of GW and which detector will be able to observe them. As we can see LISA will give access to the observation of GWs produced by massive binaries and extreme mass ratio inspirals, which cannot be detected by ground-based detectors. We can also see how ET and CE will change the GWs observation for CCSNe. With this big increase in sensitivity, the number of detections will dramatically rise in the future.

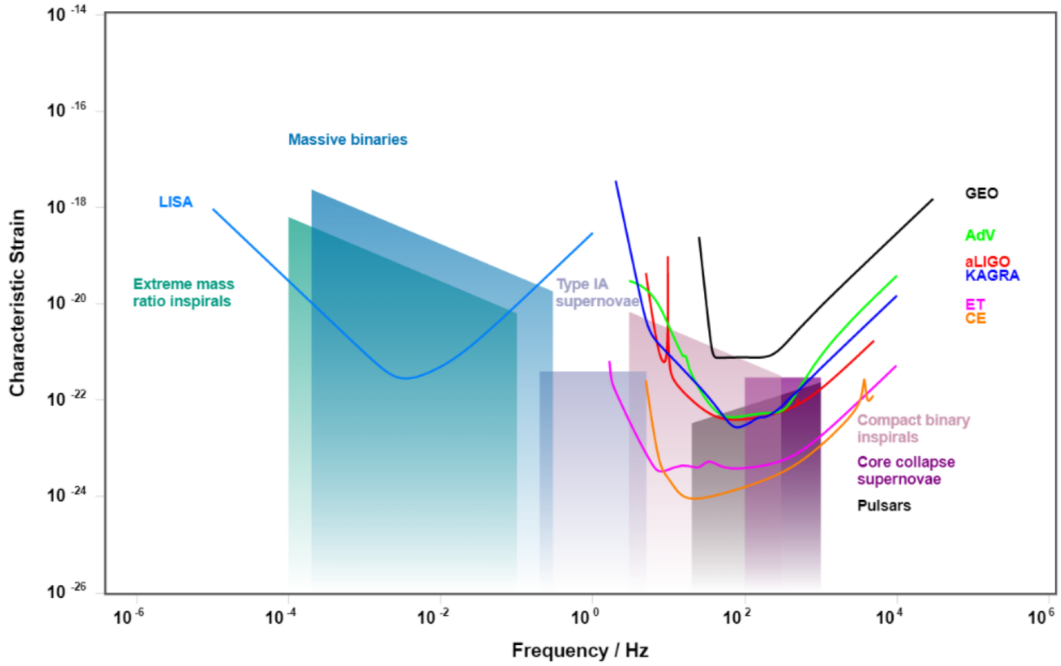


Figure 7: GWs sensitivity curves for the current detectors and future detectors and possible GWs sources to be detected. The image was produced with *http://gwplotter.com*.

4

Deep Learning

At least since ancient Greece, inventors have thought of creating an intelligent machine: a machine that could think, speak and/or perform any other action that resembles human cognition and intuition.

At first, Artificial Intelligence (AI) research was focused on solving problems that could be considered too arduous for the human mind, e.g. complex calculation problems. This was accomplished by writing long programs with several rules for the computer to follow. These abstract and formal tasks are easier for computers to follow but in regard to tasks that require some knowledge about the world, such as object recognition or speech, the subject is quite different as this knowledge is subjective and intuitive. There is a need for AI systems to acquire knowledge on their own, a capability that we call Machine Learning (ML).

The performance of ML depends on the representation of the data (knowledge) available to these algorithms. These implementations use very large datasets to extract conclusions, which are achieved through automatic self-improvement processes.

The ML algorithms can be grouped into two categories depending on the type of learning performed while training: *supervised* or *unsupervised*. In *supervised learning algorithms* the collection of examples (dataset) provided has features incorporated within, i.e. each item has attributed labels or targets. During training, the algorithm evaluates the input data and its labels and tries to find a model that can accurately predict these labels. On *unsupervised learning algorithms*, the items of the dataset do not have targets associated with them thus their goal is to learn useful structures to split the dataset into groups of similar properties (*clustering*). In this dissertation, the algorithms implemented only apply *supervised learning*.

An important implementation of ML is Deep Learning (DL). DL models aim to learn how to represent the world as a hierarchy of concepts, i.e., each concept is defined in relation to other simpler concepts. The biological brain serves as an inspiration to engineer systems of DL, where the data is evaluated by several layers of a NN, each layer with increasing abstraction [88].

There is a vast number of implementations of NNs but the main focus of this work is on Convolutional Neural Networks (CNNs), specifically Residual CNNs (ResCNNs).

4.1 Neural Networks

As a first approach, we can look at a biological neuron and trace a model for the artificial neurons. In Figure 8, we can see the schematics of a real neuron. It consists of *dendrites*, a *nucleus*, an *axon*, and *axon terminals* that allow the connections between different neurons via *synapses*. Making relation to artificial neurons, the *dendrites* will be the *input terminals*, the *nucleus* represents the *CPU*, the *axon* is the *output wire*, and the *axon terminals* will be the *output terminals* [89].

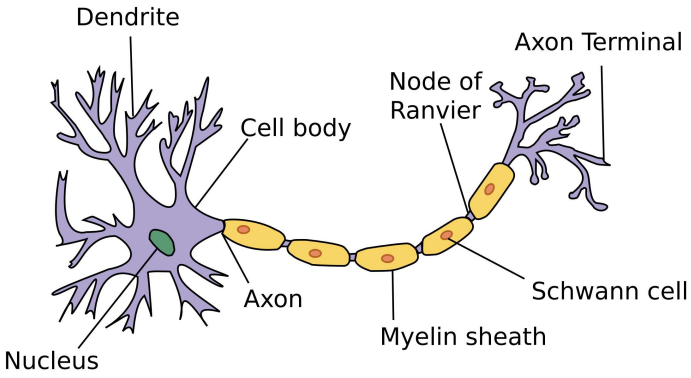


Figure 8: Representation of a biological neuron. Image from [89].

The information (x_i) will arrive at the neurons through the *dendrites* weighted by the *synaptic weights* (ω_i) determining the effect of these *inputs* ($\omega_i x_i$). These multiple weighted *inputs* will be collected at the *nucleus* as a weighted sum $\hat{y} = \sum_i \omega_i x_i + b$, where b is the *bias* (or *offset*). This information is then sent through the *axon* for further processing, $f(\hat{y})$, and is transmitted to other neurons until the information reaches its destination.

We call the artificial neuron a *perceptron* (the basic unit of a NN). An essential kind of DL model is the Multilayer Perceptron (MLP) which is composed of many layers of *perceptrons*. The NN starts with an *input layer* followed by other layers, each with a different function $f^{(i)}$. The last one will be the *output layer*. During training, these intermediate layers will not have direct access to the input labels. They are called *hidden layers*. If all neurons in a *hidden layer* are connected to all neurons in the previous and the following layers, the network is considered a Fully Connected Neural Network (FCNN). Not all architectures are constructed this way and the connectedness of the layers can vary.

The initial information provided by the *input* is going to propagate to the *hidden layers* and produce the *output* at the end of the network, in a process called *forward propagation*. For example, we can consider a composition of three functions, $f^{(1)}$, $f^{(2)}$, and $f^{(3)}$. These functions are connected to form $f(\boldsymbol{x}) = f^{(3)}(f^{(2)}(f^{(1)}(\boldsymbol{x})))$. It is the length of this chain that provides the model its *depth* (hence the name *deep learning*).

The goal of the MLPs is to approximate some function f^*. As an example, $\boldsymbol{z} = f^*(\boldsymbol{x})$ is a mapping of the *inputs* $\boldsymbol{x}$ to a certain category $\boldsymbol{z}$. The MLP will try to find a mapping $\boldsymbol{z} = f(\boldsymbol{x}; \boldsymbol{\theta})$ and learn the parameters $\boldsymbol{\theta}$ that best approximates f^*. The learning algorithm has to decide how to use each layer to

reach the best approximation of $f(\boldsymbol{x})$ for the target *output*.

Each *hidden layer* of the network is vector-valued and it is the number of elements of each vector (dimensionality) that determines the *width* of the model. These elements are the units of the NN, i.e., the *perceptrons*. Each layer contains several units working in parallel and is represented by a vector-to-scalar ***activation function***, $f^{(i)}(\boldsymbol{x})$. Each unit receives an *input* and calculates its own activation value. It is through these *activation functions* that the network decides if a neuron should be activated or not, adding non-linearity to the models.

To perform the fitting of a function we must first decide how we will measure the proximity of our prediction to the target. In order to quantify the distance between the real and the predicted values of the targets in each layer, we use ***loss functions***, $L(\boldsymbol{\omega}, b)$, where $\boldsymbol{\omega}$ is a vector composed of the weights ω_i of each unit. We want this loss to be as low as possible, where having a loss of 0 would imply a perfect prediction.

As an example, for linear regression, it is common to use the *mean squared error* to calculate the *loss function*. Considering, for a certain example i, $\hat{y}^{(i)}$ to be the predicted value and $y^{(i)}$ the true value, the *loss* would be given by $l^{(i)} = \frac{1}{2}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$. The $\frac{1}{2}$ factor makes no difference in defining the minimum but is convenient for computing the derivative of the *loss*. The *loss function* will then be the sum of the individual *losses* on the training set,

$$L(\boldsymbol{\omega}, b) = \frac{1}{n}\sum_{i=1}^{n} l^{(i)} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2}\left(\boldsymbol{\omega}^T \boldsymbol{x}^{(i)} + b - y^{(i)}\right)^2.$$

The updating of *weights* throughout the training in DL NNs is done by a method called *backpropagation* by calculating the gradient of the chosen *loss function*. This is accomplished by recursively multiplying the gradients of each layer through the chain rule for differentiation. To optimize the model, these *weights* will be updated in a way that incrementally lowers the *loss function*. The complete algorithm is generally called ***Gradient Descent***. This can be extremely slow as it implies calculating the gradient of the *loss* for every example on the dataset every time we need to update the *weights*. A variation of this is called *Minibatch Stochastic Gradient Descent*, where we do this computation for a smaller sample of the dataset, a *minibatch* β composed of a fixed number of examples. Considering a predetermined positive value η, the *weights* will be updated following the transformation

$$\boldsymbol{\omega} - \frac{\eta}{|\beta|} \bigtriangledown L \rightarrow \boldsymbol{\omega}', \tag{4.1}$$

where $|\beta|$ represents the number of examples present in the *minibatch* and η denotes the Learning Rate (LR). These two values have to be defined manually before training. They are called *hyperparameters*.

Different architectures of NNs can be implemented to achieve faster and better results. Two popular architectures will be presented in the next sections, CNNs and Residual Networks (ResNets), as they are of great relevance for the analyses of GWs *Time Series*.

4.1.1 Convolutional-Neural Networks

Convolutional Network architectures are vastly used, especially for object detection, image recognition, and semantic segmentation [90, 91]. They are notably important for the analysis of grid-like data, for example, *Time Series* data (1D) or image data (2D), and are based on the premise that nearby points are typically related to each other (for example, the nearby pixels in an image). CNNs tend to be computationally efficient, as they are easy to parallelize and require fewer parameters than FCNNs.

Instead of general matrix multiplication, these networks implement in at least one of its layers a mathematical linear operation to the *inputs* and *weights* provided, called ***Convolution***. A *Convolution* between two functions x and w can be defined as $(x * w)(t) = \int x(a)w(t-a)da$, where x will be our *input* and the *weighting function* w will be our *kernel*. For discrete objects the previous equation becomes $(x * w)(t) = \sum_a x(a)w(t-a)$. As we are going to treat multidimensional arrays of data, as well as a multidimensional array of parameters, we can write this relation as a function of a two-dimensional *input* (I) and a two-dimensional *kernel* (K) as $(I * K)(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n)$. *Convolution* is commutative, so we can also write $(I * K)(i,j) = (K * I)(i,j) = \sum_m \sum_n I(i-m, j-n)K(m,n)$ which is easier to implement computationally. This commutative property comes from flipping the *kernel* relative to the *input*. As this is not an important property for NNs, many implementations use the *Convolution* operation but without flipping the kernel, called *Cross-correlation*, which is given by

$$(I * K)(i,j) = \sum_m \sum_n I(i+m, j+n)K(m,n).$$

Usually, when implementing CNNs, we set the *kernel matrix* to be smaller than the *input*. The number of connections between m *inputs* and n *outputs* is limited, where each *output* instead of dealing with all m *inputs*, will only deal with some of them, k. This way instead of needing $m \times n$ parameters, the algorithm will only need to store $k \times n$ parameters, improving the efficiency of the NN by reducing memory and run time. We call this *sparse interaction*, opposing the idea of a FCNN. In Figure 9, we can see a representation of the *Cross-correlation* operation applied in a convolutional layer.

CNNs commonly use another operation called ***Pooling***. Similarly to what happens in the *Convolution function*, a *Pooling function* replaces the output, at a certain location, with a summary statistic of the nearby outputs. For example, the *maximum pooling* (*average Pooling*) function will report the maximum (average) *output* value within the nearby values. The use of these functions benefits the invariance of the representation to small shifts of the *input*, i.e., if we translate the *input* by a small step, most of the *pooled output* values will not change.

The architecture of a CNN can be done in many different ways and with other different operations. Still, they tend to be a collection of stacked blocks switching between the *Convolution*, *Activation*, and *Pooling functions*. A combination of these three functions is what we call a *convolutional layer* and their purpose is to extract the more relevant parts of a given dataset. CNNs will be a combination of multiple *convolutional layers* ending, however, with a *fully connected layer* to ensure the fulfillment of the task intended for the model.

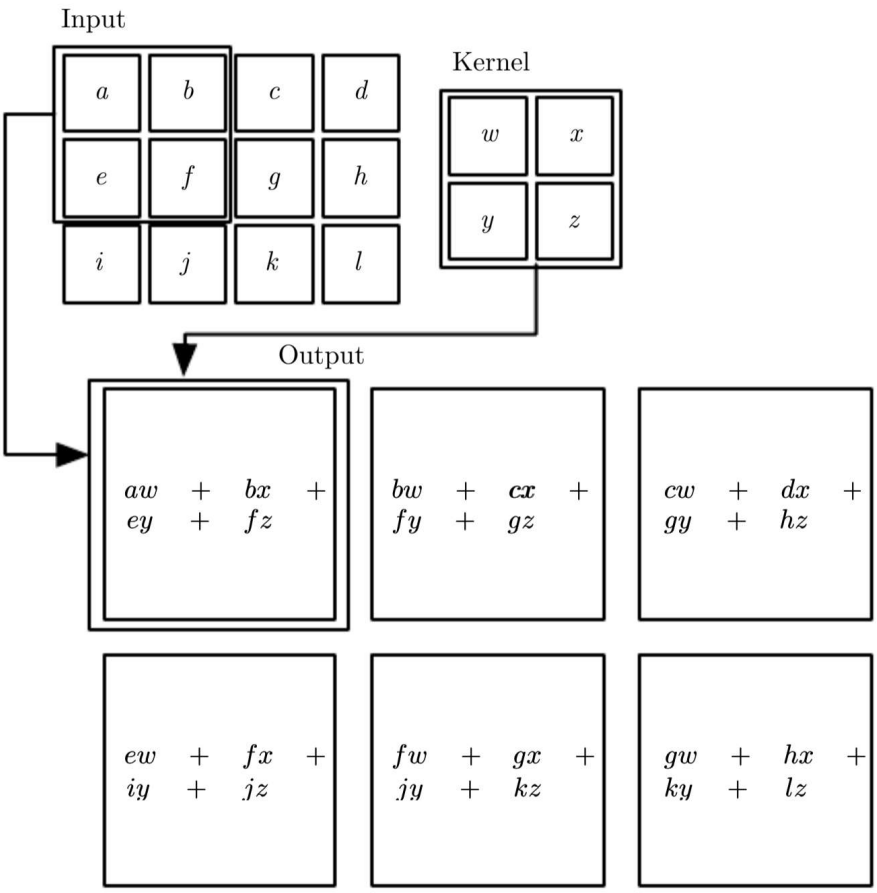


Figure 9: Representation of the *Cross-Correlation* operation in a convolutional layer.

4.1.2 Residual Blocks

Increasing the *depth* of a network can lead to a saturation of accuracy and even its degradation. Having more layers can be associated with higher training errors which can be prevented with the addition of ***Residual Blocks*** [92].

The main idea of the *Residual Blocks* is to explicitly let some layers fit a residual mapping. Having an *input* x and the desired mapping $f(x)$, the residual mapping will be given by $h(x) = f(x) - x$. It is hypothesized that optimizing the residual mapping is easier than optimizing the original (unreferenced) mapping. Considering the identity mapping, where $f(x) = x$, it would be easier to drive the *weights* and *biases* to zero to get $h(x) = 0$ than to actually drive $f(x)$ to x. *Residual Networks* take advantage of this process by performing identity mapping with shortcut connections, i.e., the identity mapping is applied, skipping one or more layers, and their outputs are added to the outputs of the stacked layers. Figure 10 shows an example of a residual block and how the identity mapping is applied.

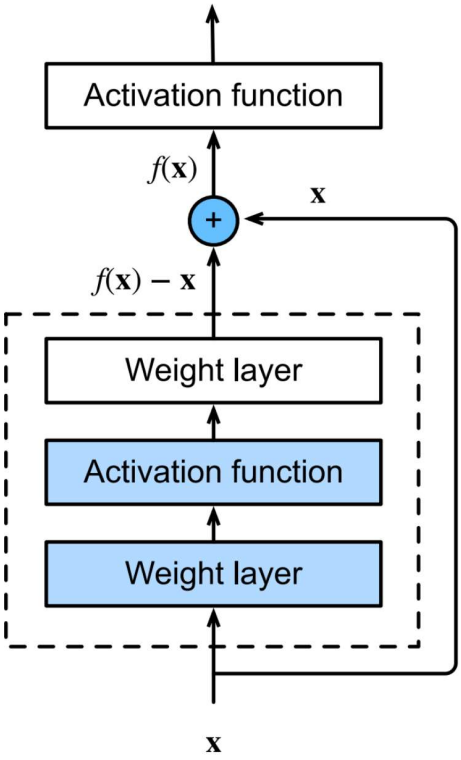


Figure 10: Representation of a *Residual Block* by applying identity mapping.

4.2 Regularization

When using ML algorithms we need to make sure good models perform well not only with the initial training dataset but also with any new dataset. *Regularization* is the collection of actions designed to reduce the *test loss* even if at the expense of an increased *training loss*. Two challenges that *regularization* methods try to overcome are ***underfitting*** and ***overfitting***. *Underfitting* arises when the algorithm does not find a model with a sufficiently low *training loss* value. *Overfitting* is the case mentioned above when the *training loss* is smaller than the *test loss* and with further training, these two values tend to diverge even more.

Changing the *capacity* of the model can help with these challenges. The *capacity* of a model is its ability to fit different functions. Low *capacity* can lead to *underfitting* while high *capacity* can lead to *overfitting*. The *capacity* of an ML algorithm can be controlled by choosing its hypothesis space, i.e., the set of functions the algorithm can choose from, to approximate a solution. In Figure 11, we can see an example of three different hypothesis spaces that lead to different results.

Instead of just changing the model's *capacity*, the learning algorithm can also be given preferences on its options, i.e., the algorithm will have several options available but some of them will be preferred to others (different weights). This way, an approximation made with a less preferable function will only be picked if its performance is significantly better than an approximation made with a more desirable function.

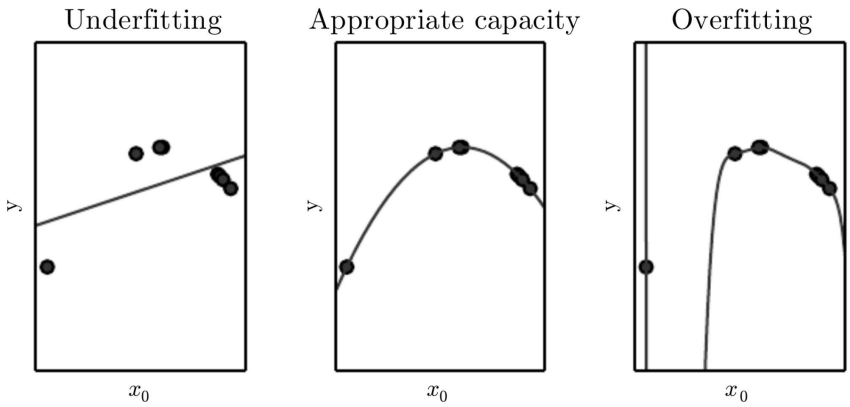


Figure 11: Representation of three different models with the same training set. (Left) A linear function was fit to the data leading to *underfitting*. (Center) A quadratic function was fit to the data with no significant *overfitting* or *underfitting*. (Right) A polynomial of degree 9 function was fit to the data leading to *overfitting*.

The inclusion of ***Weight Decay*** (also known as L^2 parameter norm penalty) can be one way of modifying the training criterion, a function $J(\boldsymbol{\omega})$ that expresses the *weights* of preference. The goal is to minimize the sum of the *mean squared error* on the training and

$$J(\boldsymbol{\omega}) = MSE_{train} + \lambda \boldsymbol{\omega}^T \boldsymbol{\omega},$$

where λ is a value set from the beginning of the training that controls the preference for small weights. When $\lambda = 0$, we do not have a preference, and for large λ, the *weights* are forced to be smaller. By minimizing $J(\boldsymbol{\omega})$, we find the *weights* that provide the best balance between being lower and being able to fit the training data. The values for the *weight decay* have to be chosen carefully. When λ is too low (almost 0) it can lead to *overfitting*, while a large λ may lead to *underfitting* [88].

Another important regularization approach is ***Early Stopping***. At some point in the training, the *training loss* may continue to decrease steadily while the *validation loss* starts to rise. To prevent this, we can store the model's parameters every time the validation error decreases. This allows us to return to the parameter settings at the point of lowest validation error whenever the model starts to increase this error. When the training process terminates, rather than the last model obtained, we can retrieve the model with the set of parameters corresponding to the lowest *validation error*.

Dropout is also a powerful method that can be used for *regularization*. It provides a computationally inexpensive approximation to train many NNs and average the predictions of the models. Its goal is to increase the stability and accuracy of the methods against small perturbations that can happen during training, introducing uncertainty to the model. *Dropout* consists in training an ensemble of all the sub-networks that could be formed by removing units from the internal layers of the base network. As an example, this can be accomplished by simply setting the *output* value of such units to zero during training. For faster performance, instead of training each sub-network on the full dataset, we can do this training in *minibatches*. The variation of *Dropout* used in this work is ***Monte Carlo Dropout*** (MC Dropout). It considers *Dropout* training in NNs an approximation of *Bayesian Inference* in *Gaussian* processes [93]. In

this case, we will only apply *Dropout* after training, in the test phase, getting several predictions for each model that we can average and analyze.

There are many other types of *regularization* methods for DL besides *weight decay*, *early stopping*, and *Dropout*, but those fall out of the scope of this dissertation and for that reason they will not be discussed.

4.2.1 Cyclical Learning Rates

Setting the right *hyperparameters* is essential to reduce training time and to improve the performance of the methods used. The right choice for the LR, *batch size*, and *weight decay* can be decisive to avoid *underfitting* or *overfitting*. We have already seen how *weight decay* can influence training. In this section, we are going to describe a method to find an optimal LR and how we can make this LR vary during the training to improve the performance of the network.

The LR helps to control how much the model changes when we perform the update of the *weights* (eq. 4.1). When the value of LR used is too small, the network is prone to *overfitting*, and for values of LR too large the training will tend to diverge. Leslie Smith proposed in his paper [94] a method to improve the performance of a network, depending on the use of LR during training. This proposal consists of using **Cyclical Learning Rates (CLR)** instead of having the value of LR fixed during the full training. This method lets the LR value vary between a minimum and a maximum value (boundaries). A cycle of the CLR method involves two steps. The first one lets the values of LR increase linearly from the minimum to the maximum. The second step forces the LR to decrease linearly. The length of each step (stepsize) is the number of iterations, usually called *epochs*, which must be set before training.

With this method, instead of finding an optimal value of LR, which can be very difficult, we need to determine an interval of LR, knowing that the optimal value will belong to this interval. In Figure 12, we can see that this approach leads to similar accuracy when compared to a fixed LR, with a minimal increase, but making the training much faster.

In addition, Smith also developed a method to help find an optimal interval for the LR, called the *LR range test*. This is a simple test performed before the actual training takes place to understand how the LR influences the accuracy of the used algorithms for a given dataset. To accomplish this, a single training of several *epochs* is done, starting with a small LR and increasing linearly its value. An example of the resulting plot is shown in Figure 13. The LR of the extreme of the plot (where the accuracy is higher) should be considered as the maximum boundary of LR. The minimum boundary should be the value of LR where the accuracy starts to increase.

In a second paper [95], Smith and Topin introduced the concept of *super-convergence* where they show that, in certain conditions, the training of NNs can be much faster, as we can see in Figure 14. The technique to achieve *super-convergence* is to do CLR with only one cycle, starting with a low value of LR and letting it increase until the maximum value and then letting it decrease again. After this cycle is complete, the training should continue decreasing the LR value by several orders of magnitude. They call this method the *1cycle policy*.

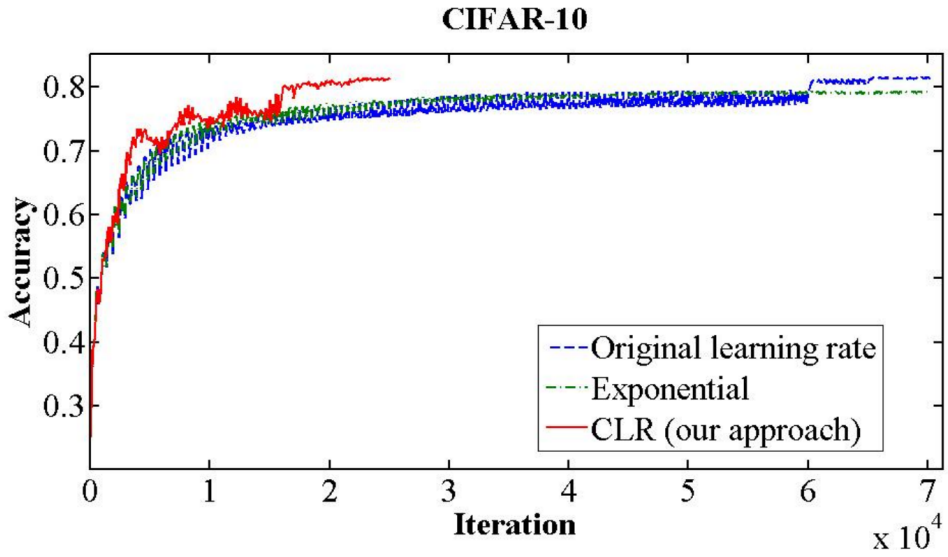


Figure 12: Results of the classification of the dataset CIFAR-10, for different types of training: Original LR, Exponential, and CLR. The plot shows the evolution of the accuracy of the different methods. These tests were performed by Leslie Smith in [94].

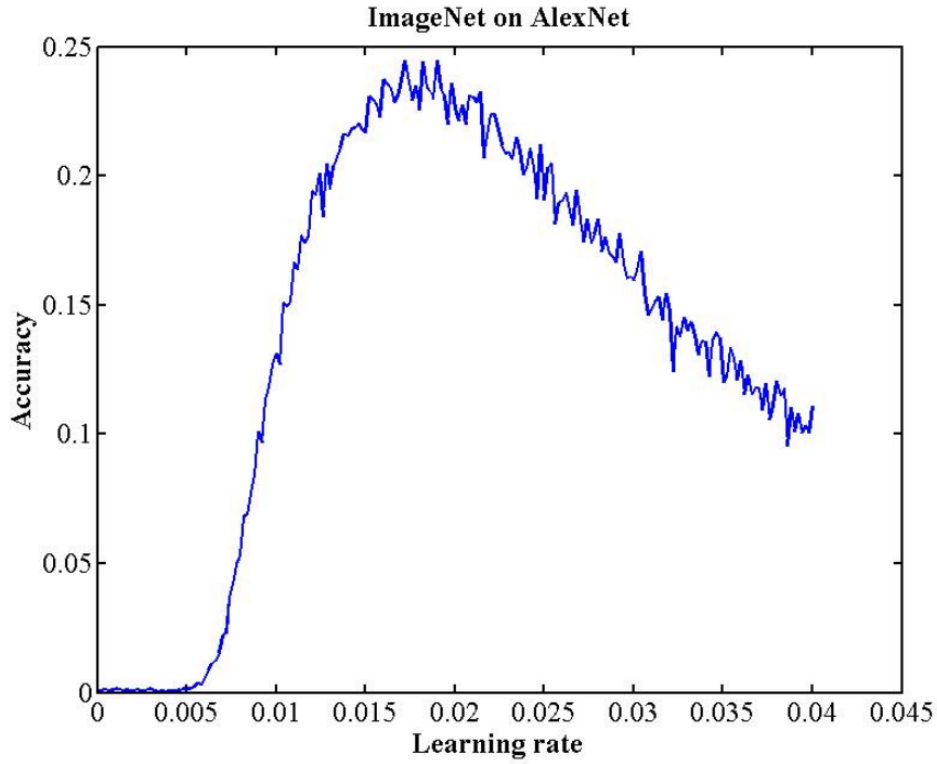


Figure 13: Example of the *LR range test*, performed by Leslie Smith in [94]. These results were obtained for a *classification* test with an *AlexNet* network.

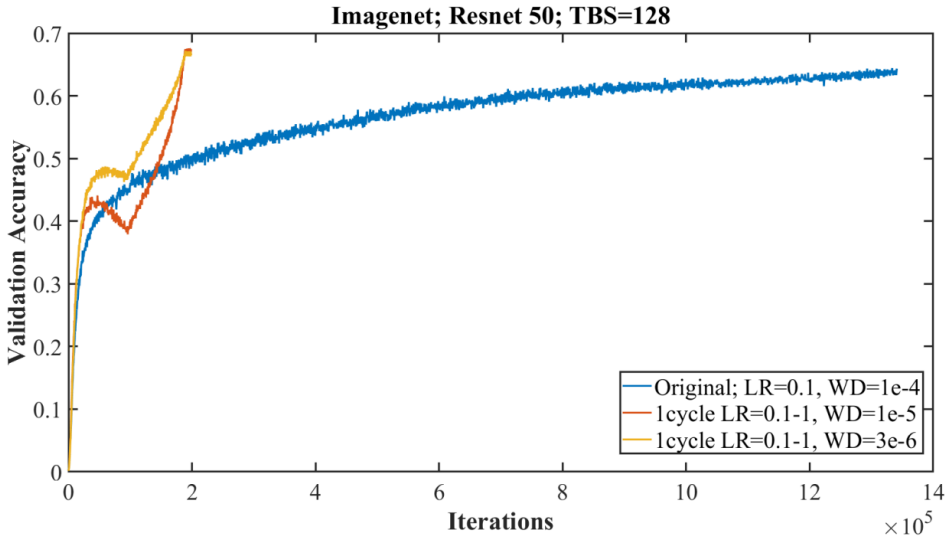


Figure 14: Results obtained by Smith and Topin in [95] by training a *Resnet-50* with the standard LR policy and with the *1cycle* policy.

5

Datasets and Methods

The continued development of GWs interferometer detectors has reached a point where their sensitivity appears to be high enough to detect the first GW signal from a CCSN, especially in the case of a Galactic event [96]. The intent of this work is to develop DL methods that could help identify and analyze these signals for the particular case of rapidly rotating CCSNe progenitors. To accomplish this, we generate datasets of *Time Series* representations of GWs and train ResCNNs to perform *classification* and *inference*. In this chapter, we will describe all the methods used to accomplish our goals.

5.1 Dataset Generation

We start this chapter by describing how the datasets used in the analysis are generated. The waveforms used in this dissertation are obtained from a catalog of CCSNe numerical waveforms developed by Richers *et al.* [37], which is publicly available in [97]. The development of this catalog, comprising 1824 signals, entails to perform an equal number of axisymmetric, numerical-relativity simulations.

5.1.1 Core-Collapse Supernovae Waveforms

To generate these waveforms, Richers *et al.* use 18 different EOS and their simulations cover a parameter space of 98 different *rotation profiles*. Each EOS applies different models and interactions. For CCSNe events, the two most extensively used EOS are those of Lattimer and Swesty (LS) [98], and H. Shen *et al.* (HShen) [99–101]. LS applies the *compressible liquid drop* model while HShen uses a *relativistic mean field* model. Many other EOS have been developed for CCSNe simulations. Besides LS and HShen, the EOS of Hempel *et al.* (HS) [102, 103], Steiner *et al.* (SFH) [104], Banik *et al.* (BHB) [105], and G. Sehn *et al.* (GShen) [106, 107] are also used in the Richers catalog. Each is based on one of the two models mentioned with different interactions or parameter sets. A summary of the EOS applied in this catalog is reported in Table I of [37].

As shown in [37], the GW signal at *bounce* is vastly independent of the EOS and especially sensitive to the ratio of *rotational-to-gravitational energy*, $T/|W|$, and to the *degree of differential rotation*, for high

rotation rates. They also show that the maximum GW *frequency during postbounce* ($f_{\text{peak}} \sim 600-1000$Hz) has a strong dependency with the EOS, being proportional to the square root of the average central rest-mass density $\sqrt{\bar{\rho}_c}$, as we can see in Figure 15. In addition, for $T/|W| \leq 0.06$ there is a linear relation between the *bounce strain amplitude* Δh and $T/|W|$, as can be seen in Figure 16. Δh is the difference between the maximum and minimum value of h before the end of the *bounce* (t_{be}).

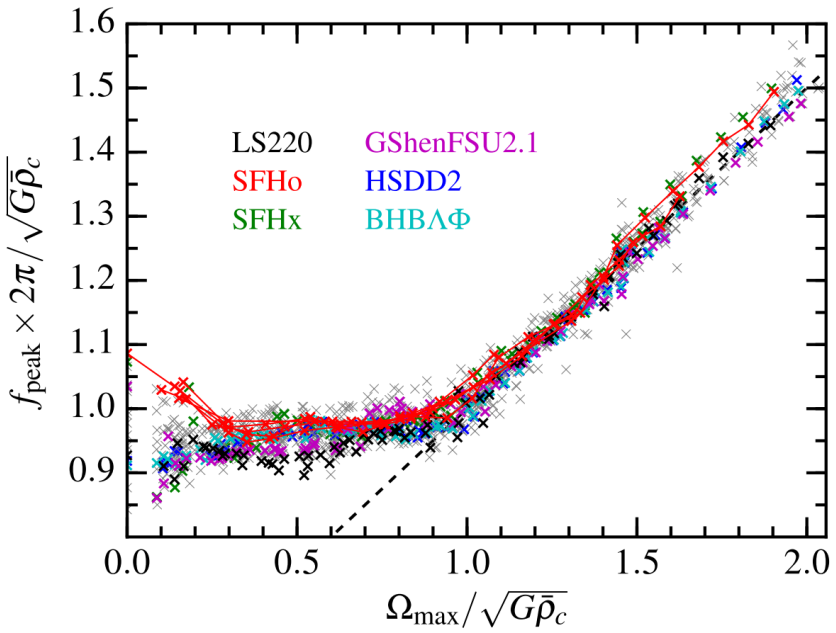


Figure 15: Values of normalized f_{peak} as a function of normalized *maximum rotation rate* Ω_{max}. The dashed line represents the relation $2\pi f_{\text{peak}}/\sqrt{G\bar{\rho}_c} = 0.5(1 + \Omega_{\text{max}}/\sqrt{G\bar{\rho}_c})$. This figure was taken from [37] (Fig. 9).

Essentially, the waveforms generated by Richers *et al.* [37] were obtained in two steps: a 1D simulation of the *collapse-phase deleptonization* with the code `GR1D` [108] and a 2D core-collapse simulation with the `CoCoNuT Code` [38]. To produce the *collapse-phase deleptonization*, Richers *et al.* used an approximation proposed by *Liebendörfer* [109] which consists of parameterizing the *electron fraction*, Y_e, as a function of density, ρ, only from spherically symmetric (1D) non-rotating GR hydrodynamic simulations, as allowing a large parameter study of axisymmetric (2D) simulations is computationally expensive. They also use this approximation for the rotating case because *electron capture* and *neutrino* interactions with matter are local and depend on the density in the collapse phase, such that the rotational flattening of the collapsing core can be considered relatively small.

Using the open-source code `GR1D`, Richers *et al.* ran spherically symmetric GR radiation hydrodynamic CCSNe simulations for a 12 $M_\odot$ non-rotating progenitor, once for each of the 18 EOS they considered. They used `NuLib` [108] to calculate the *neutrino interaction rates*, which carry *absorption* and *emission* from nucleons and nuclei, including *neutrino blocking factors*, *elastic scattering* off nucleons and nuclei, and *inelastic scattering* of electrons. They did not take into consideration *bremsstrahlung*[1] and *neutrino*

[1] *Bremsstrahlung*, meaning "braking radiation", is the radiation emitted by free electrons subjected to electric fields.

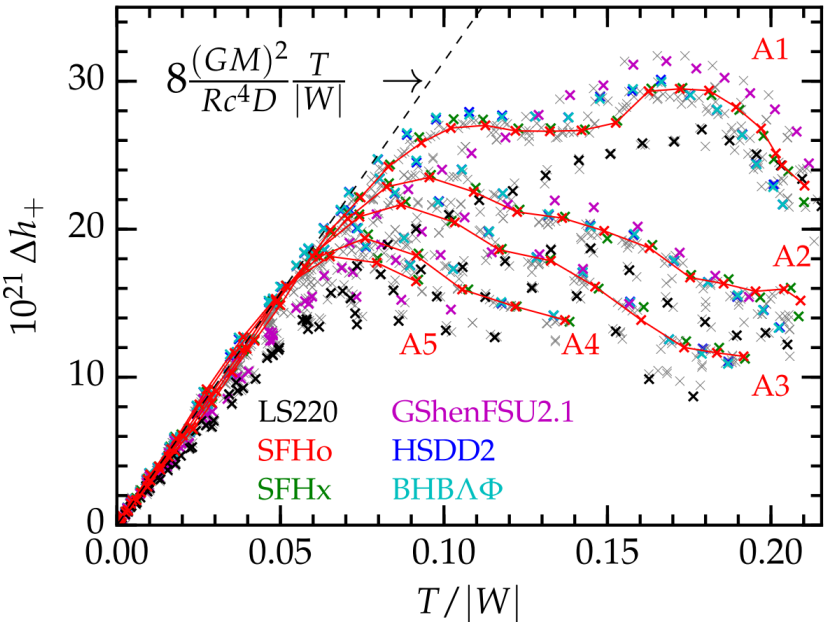


Figure 16: Values of Δh as a function of $T/|W|$, showing a linear relation for $T/|W| \leq 0.06$. This figure was taken from [37] (Fig. 6).

pair creation and annihilation, as they lack importance during the collapse and after core bounce. To finally obtain the $Y_e(\rho)$ parameterizations, at the time when the central Y_e is at minimum (nearly at core bounce), Richers *et al.* registered the state of the simulation and created a list of the Y_e and ρ values at each radius. To be certain that simulations will not find a density outside the range provided in the curves, they extrapolated them to high densities with a constant Y_e value.

In a second step, using the `CoCoNut` code, they performed axisymmetric (2D) CCSNe simulations assuming the CFC approximation for the spacetime metric. They first generated the initial conditions for the simulations by imposing a *rotation profile* on the pre-collapse star

$$\Omega(\omega) = \Omega_0 \left[1 + \left(\frac{\omega}{A}\right)^2\right]^{-1}, \tag{5.1}$$

where A is a measure of the *degree of differential rotation*, Ω_0 is the maximum initial *rotation rate*, and ω is the distance from the axis of rotation. They generated a total of 98 *rotation profiles* with different parameters that have a range of *rotation rates* slow enough to allow the collapse of the star (see Table II of [37]). These 98 rotation profiles are simulated for each of the 18 EOS, resulting in a total of 1864 simulations. Those simulations were performed with the `CoCoNuT Code`, solving the GR hydrodynamic equations coupled to the CFC equations (presented, respectively in Section 2.2.3.1 and Section 2.2.3.2) on a spherical-polar mesh, assuming reflection symmetry at the equator. In those simulations, the effect of *neutrinos* on the collapse phase is treated with the parameterizations of Y_e from before and after the *bounce*. They applied the *neutrino leakage scheme* to account for *neutrino* heating, cooling, and *deleptonization*, even though this has a very small effect on the GW signal of the *bounce* and early *post-bounce*.

Finally, the gravitational waveforms are computed using the *quadrupole formula*

$$h_+ \approx \frac{2G}{c^4 D}\ddot{I} \tag{5.2}$$

where D is the distance to the source and I is the *mass quadrupole moment* (a single dot indicates one-time derivative). Figure 17 shows how a subsample of 18 waveforms vary with the 18 different EOS used in [37], for the same *rotation profile*.

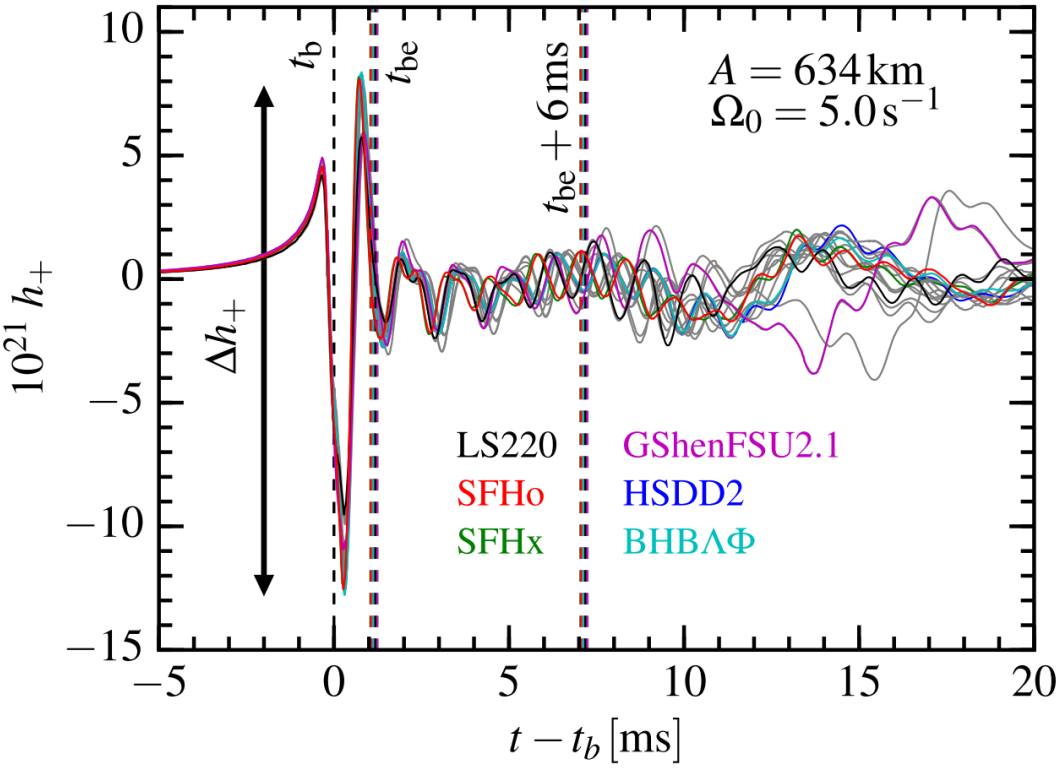


Figure 17: 18 waveforms for different EOS. The waveforms are plotted in the time domain for fixed values of the rotational profile parameters, $A = 634\mathrm{km}$ and $\Omega = 5.0\mathrm{rads}^{-1}$. Taken from Richers *et al.* [37].

5.1.2 Generating Time-Series Data

For the analysis of this work, we only used some selected waveforms from the Richers *et al.* catalog [37]. First, we remove from the list of available waveforms the simulations that did not produce a collapse within the first second of simulation time (see Table III of Richers *et al.*). We also did a selection on the *rotation profiles*, only choosing the ones that had a maximum initial (pre-collapse) *rotation rate* Ω_0 equal to or higher than 3.0 rads^{-1}. This last restriction was made as a consequence of the numerical noise present in models with slow *rotation rates*, which amplifies when performing the normalization of the signals. This effect is due to the calculation of the GW signal through the *quadrupole formula*, which is not significant for $\Omega_0 \geq 3.0\mathrm{rads}^{-1}$.

The waveforms in the catalog assume an optimally oriented source. Therefore, for an axisymmetric system only the *plus polarization* component is different from zero:

$$h_+ = h_+^{opt} \sin^2\theta \quad and \quad h_\times = 0\,, \tag{5.3}$$

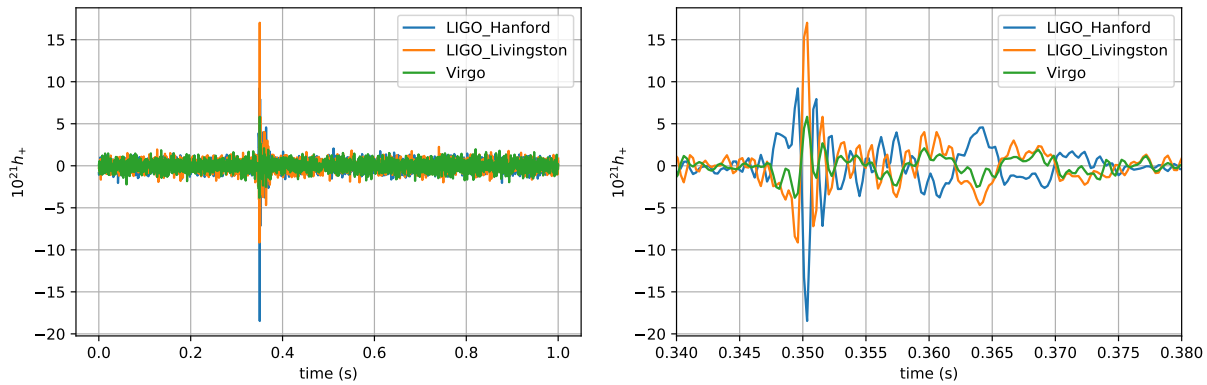
where θ is the *inclination angle*. As we set this angle to be $\pi/2$ rad we will simply have $h_+ = h_+^{\text{opt}}$. Each simulation provides four physical parameters: the strain in units of distance (cm), $D \cdot h_+^{\text{opt}}$, the timestamps, $(t - t_{\text{b}})$ (considering the time of core *bounce* as the origin), f_{peak}, and the values to calculate the Δh.

With this set of values and defining the distance of the source between 5 - 20 kpcs (i.e. a CCSN that occurs in our Galaxy), we do a *down-sampling* of the waveforms to match the sample of noise available. For the background noise, we use publicly available data from both LIGO (Hanford and Livingston) and Virgo, with an initial Global Positioning System (GPS) time $t_{\text{GPS}} = 1253326755$s, which corresponds to the end of September 2019, i.e. around the middle of the O3. We define a random time, $t_0 \in [6, 206]$s, and select a segment of the background noise with a window $[t_0 - 5, t_0 + 10]$s for the *sample noise*. Each waveform is projected into each one of the detectors with random values for the angles defining the sky position (*declination*, *right ascension*, and *polarization*). Then, we inject the projected waves into the *noise samples* for each detector, having the time of *bounce* of the signal corresponding to a random $t_{\text{b}} \in [t_0, t_0 + 0.8]s$.

We perform *whitening* on the resulting strains through *inverse spectrum truncation* using the Amplitude Spectral Density (ASD) of the noise, i.e., we transform the signal to the frequency domain by doing the *Fourier* transform, divide it by the ASD, and transform it again to the time domain. After this, we apply a *band-pass filter* from 20 Hz to 1000 Hz and *notch filters* at the individual frequencies 60 Hz, 120 Hz, and 240 Hz, for the two LIGO strains, and 50 Hz, 100 Hz, and 200 Hz for the Virgo strain. To finish, we do a final crop on the strains so we only have a window of 1s, where t_b can be between the 0.1s and 0.9s of this window.

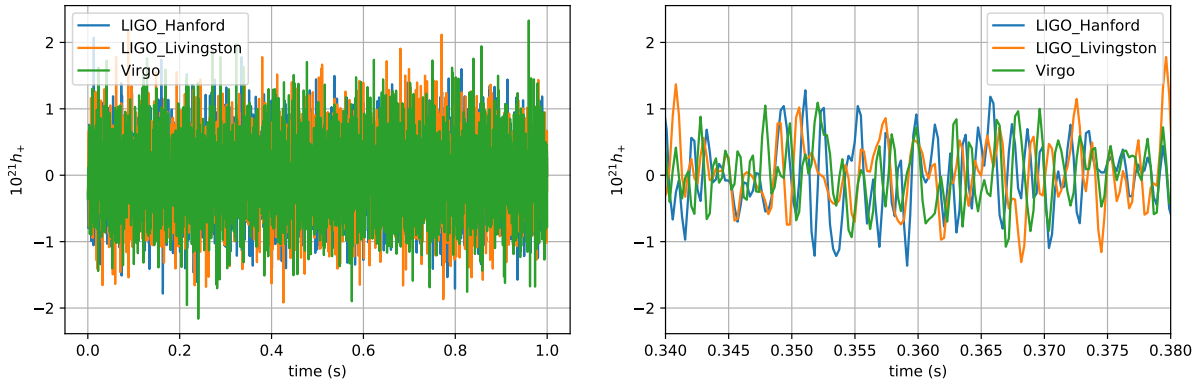
Each element of the dataset is composed of three *Time Series*, one for each detector. An example of these *Time Series* is given in Figure 18. We also needed strains with only noise for the datasets used for *classification*. When generating the strains of noise we performed the same treatment described before, only letting out the injection of the signal waveforms. In Figure 19, is represented an element with only background noise from O3.

All these processes are done in `python` [110] with specific functions from the libraries `SciPy` [111], `PyCBC` [112], and `GWpy` [113]. We also use the libraries `fast.ai` [114] and `tsai` [115] for the *classification* and *inference* tests, as will be described in the next section.



(a) (b)

Figure 18: Example of a dataset's element with an injected signal in the background noise from O3.



(a) (b)

Figure 19: Example of a dataset's element of only background noise from O3.

5.2 Neural Networks

As presented in Chapter 4, we choose ResCNNs to test the DL methods on CCSNe GWs signals. ResCNN is an implementation of CNN with *residual learning blocks*. This network, proposed in [116], has 6 *convolutional layers* and uses *batch normalization* techniques and different *activation functions*. The most used *activation function* is *Rectified Linear Unit* (ReLU) which is a kind of linear function that returns the value of the input in case it is positive, otherwise, it will return zero. Other versions of this function are also used: *Leaky ReLU* (LReLU), *Parametric ReLU* (PReLU), and *Exponential Linear Unit* (ELU). It is desirable to use different *activation functions* in different layers to achieve better performance.

As the analysis of the *Time Series* needs the precise location of the features, the *pooling* operation was removed. Another reason for this choice is the fact that *pooling* reduces the number of *inputs* given to the next layer, limiting the information available.

For this architecture, the usual *Fully Connected Layer* at the end of the network was replaced by a *global average pooling* and a final layer performing *Softmax* as its *activation function*. *Softmax* computes the exponential of the values and normalizes the value according to

$$f(z^{(i)}) = \frac{e^{z^{(i)}}}{\sum_{j=1}^{n} e^{z^{(j)}}}, \tag{5.4}$$

where n is the number of elements that compose the vector z given by the previous layer. Doing this in the last layer makes the output of the network a probability distribution of its own predictions.

The architecture of these networks is represented in Figure 20.

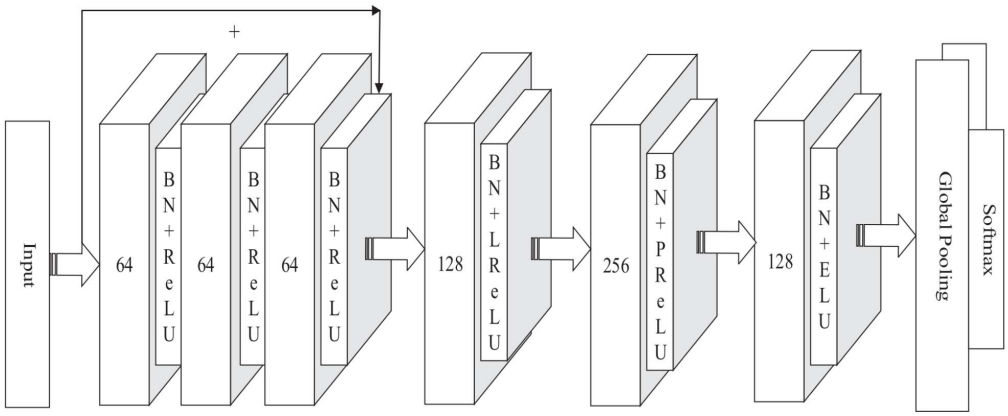


Figure 20: Representation of the ResCNN's architecture.

We use the function `fit_one_cycle` to perform the training of the network, defining the maximum LR and the *weight decay*. This function of `fastai` internally uses the usual `fit` function but applies the *1cycle policy* for better results, known as *super-convergence*, which is explained in Section 4.2.1. To train the models, we divide the dataset into a *training set* and a *validation set*. For *classification*, we use a split of 80% for training and 20% for validation while for parameter estimation we choose 70% for training and 30% for validation.

The maximum LR and *weight decay* used should be adapted to each dataset and architecture [117]. In our case, to find the best values for our datasets we use another `fastai` function called `lr_find` based on the LR range test [94]. Given a range for the LR, this function trains the network for some iterations starting on a minimum LR value and increasing it throughout the training until it reaches a maximum value. For each iteration the *loss* is recorded and, at the end of the training, the function plots the *losses* as a function of this LR values. The suggested value for the maximum LR used on the training with `fit_one_cycle` will be the *steepest point* of this plot, i.e., the LR value where the *loss* decreases more rapidly (learns faster). With some tests, we verified that for our datasets the best results were for *weight decays* of 10^{-3} and maximum LR on the order of 10^{-3}.

To calculate the *accuracy* (Acc) of the model, used in the *classification* method, we used the formula:

$$Acc = \frac{\text{True_Positives} + \text{False_Negatives}}{\text{True_Positives} + \text{True_Negatives} + \text{False_Positives} + \text{False_Negatives}}, \tag{5.5}$$

where the *True Positives* (*True Negatives*) are the correctly predicted signals (noise) and the *False Positives* (*False Negatives*) are the items predicted as signals (noise) but labeled as noise (signal). We also plotted the Receiver Operating Characteristic (ROC) curve for the *classification* networks. This is an illustration of the performance of the *classifier* as its threshold changes. The plot provides the *True Positive Rate* and *False Positive Rate* for different values of the threshold. The *False Positive Rate* is calculated with:

$$FPR = \frac{\text{False_Positives}}{\text{False_Positives} + \text{True_Negatives}}, \tag{5.6}$$

and the *True Positive Rate* with:

$$TPR = \frac{\text{True_Positives}}{\text{True_Positives} + \text{False_Negatives}}. \tag{5.7}$$

This network is considered to have a perfect performance when the Area Under the Curve (AUC) for this plot is equal to 1.

For the *regression* method, before giving the dataset to the network we perform a *standard normalization* of the parameters of the signals. To find the predictions of the network in both methods, *classification* and *regression*, we use *MC Dropout*, passing each signal through the network 100 times. With this, each event prediction will be the *mean value* of the 100 evaluations.

Finally, we used Principal Component Analysis (PCA) to evaluate the prediction distributions for the *regression* tests. We use directly the `python` function of PCA from the `scikit-learn` library [118]. PCA can be used for *dimensionality reduction*, being able to identify the most relevant information (features) in datasets, but in this work, we only use PCA to perform *Singular Value Decomposition* on the two variable distributions we already had (predicted values and real values). With this, we got the two orthogonal eigenvectors to calculate the slope and with the singular values, i.e, $\sqrt{\text{eigenvalues}}$, we calculated the variation, σ, of the distribution.

The 100k element dataset was created and trained on `ARTEMISA` (*ARTificial Environment for Machine Learning and Innovation in Scientific Advanced Computing*) an artificial environment providing CPU and GPU power based on the Universitat de València [119].

6

Results and Discussion

Two different methods of DL are applied to carry out the analyses of GWs signals from CCSNe: *classification* and *regression*. All the results from the performance of these methods are presented and discussed in this chapter.

6.1 Classification Results

A common test performed by DL methods is *classification*. This method is used in this work to distinguish strains of background noise from other strains with injected GWs signals. All strains of noise are segments of real noise from the aLIGO and AdV detectors on the first half of O3 (O3a).

6.1.1 10k Dataset

As a first test, a dataset of 10000 *Time Series* was used: 5000 with only detector noise and another 5000 with injected CCSNe signals. The signals for this dataset were generated with the conditions reported in Table 1.

Parameter	Range
SNR	> 5
Distance	$[5, 20]$ kpc
Inclination	$\pi/2$ rad
Declination	$[-\pi, \pi]$ rad
Right Ascension	$[0, 2\pi]$ rad
Polarization	$[0, 2\pi]$ rad

Table 1: Range of parameters used to generate the datasets for classification and regression.

The training is repeated for different numbers of *epochs* to study the model's evolution. The maximum LR was set to 3×10^{-3} and the *weight decay* as 10^{-3} to perform the fits. It should be noticed that in all these tests the number of *False Positives* obtained was always zero with a 0.5 threshold.

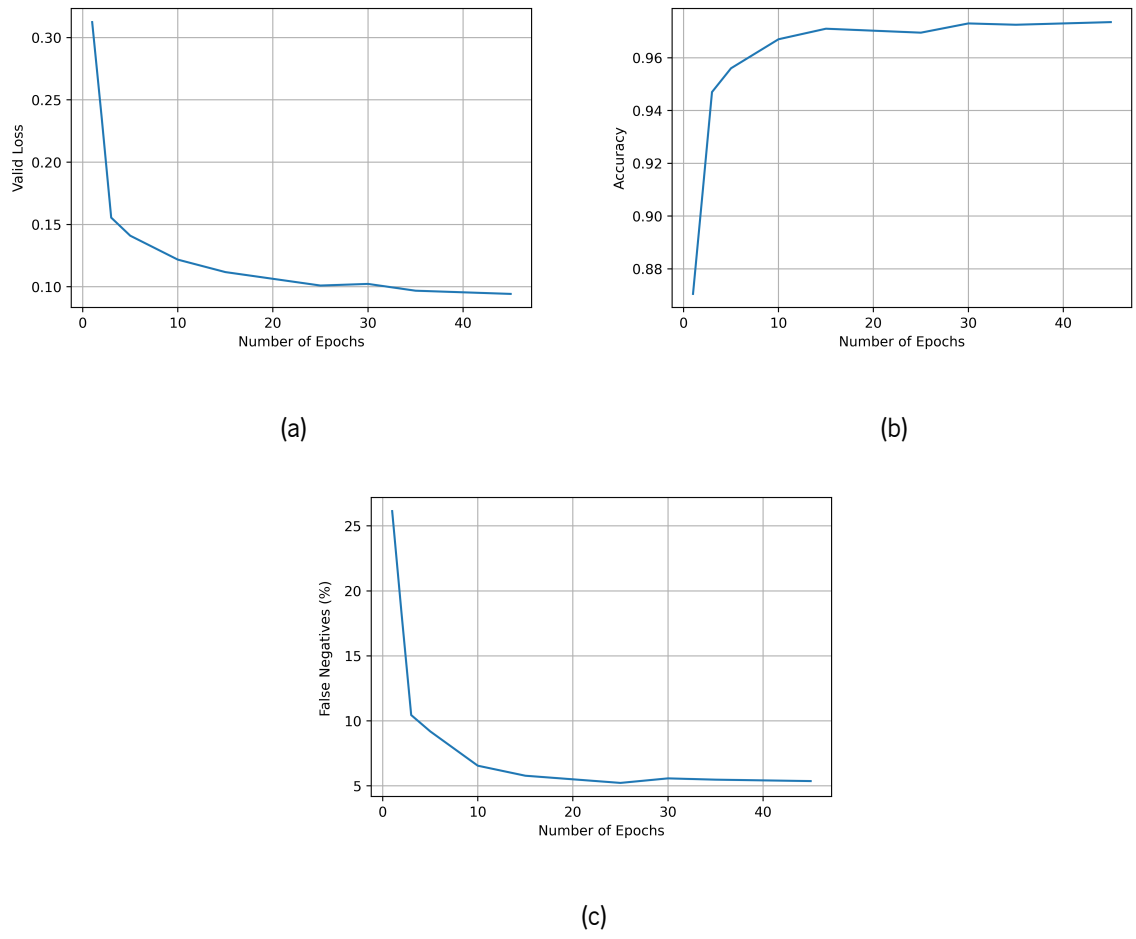


Figure 21: Evolution of the values of *Valid Loss* (a), *Accuracy* (b), and percentage of *False Negatives* (c) obtained when performing different *classification* tests with a varying number of *epochs*.

Figure 21 depicts the evolution of the *Valid Loss*, *Accuracy*, and *percentage of False Negatives* (number of False Negatives per number of actual signals) as a function of the number of *epochs* used for the different tests. After around 20 *epochs* these values reach a *plateau*.

As an example, when the number of *epochs* is set to 25, the network finds the minimum *Valid Loss* to be 0.10, after 23 *epochs*. With an *Accuracy* of 97% we get the *confusion matrix* shown in Figure 22. Only 51 real signals are predicted as noise, which is only 5% of all real signals. Correspondingly, Figure 23 displays the ROC curve of this *classification* test, providing a 0.99 AUC.

To better understand the results of the misclassified 51 signals of Figure 22, Figure 24 shows the distribution of the scores obtained, where each point represents an element of the dataset labeled as actual signal but predicted as noise (*False Negatives*). In these plots, we depict our three key CCSNe parameters, *distance*, Δh, and f_{peak}, as a function of the Signal-to-Noise Ratio (SNR) with different colors for the scores given by the model. All the misclassified signals have SNR below 10, which is expected. In those cases, it is difficult to distinguish the signal from the random peaks of the background noise as the amplitudes are similar.

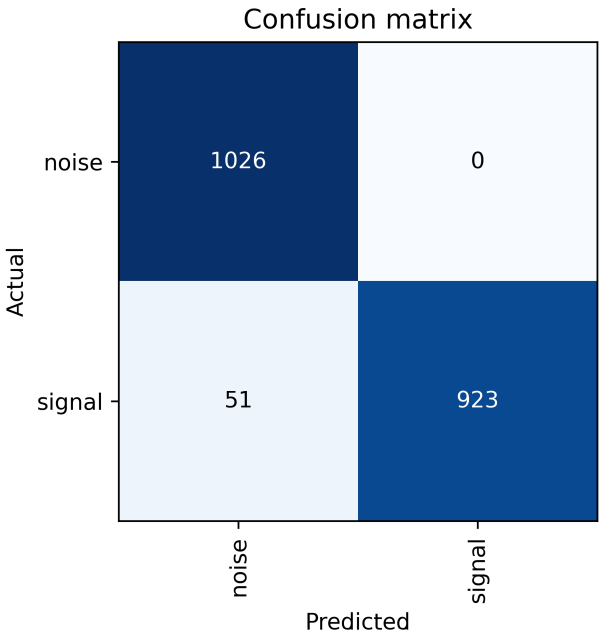


Figure 22: Confusion Matrix for the *classification* test with 25 *epochs* with a dataset of 10k *Time Series*, giving the number of strains predicted as noise or as signal versus their actual labels.

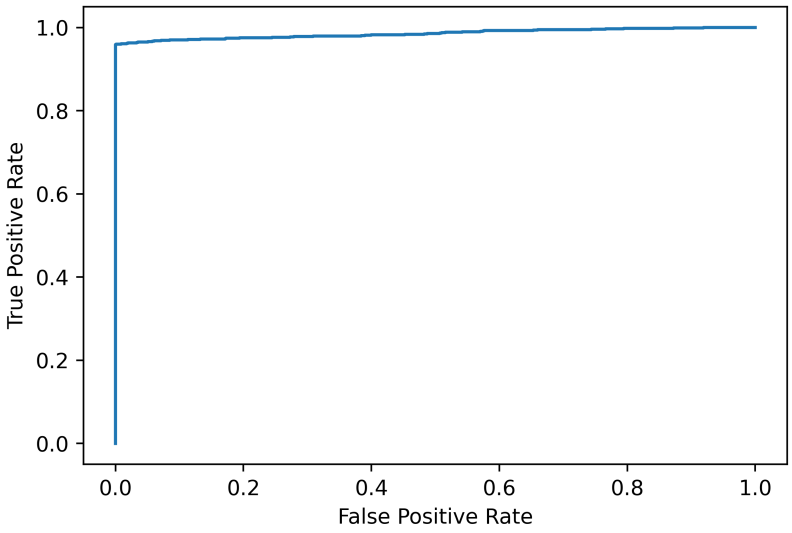


Figure 23: ROC Curve of the *classification* test with 25 *epochs* with a dataset of 10k *Time Series*.

6.1.2 100k Dataset

To confirm the previous results, identical tests were done with a ten times bigger size dataset. Similarly to the previous tests the evolution of the model is studied for trainings with different numbers of *epochs*, reaching now the *plateau* around 10 *epochs* for this new dataset. With a maximum LR of 2×10^{-3} and the same *weight decay* as before, 10^{-3}, the model yields a minimum *Valid Loss* of 0.077 and an *Accuracy* of

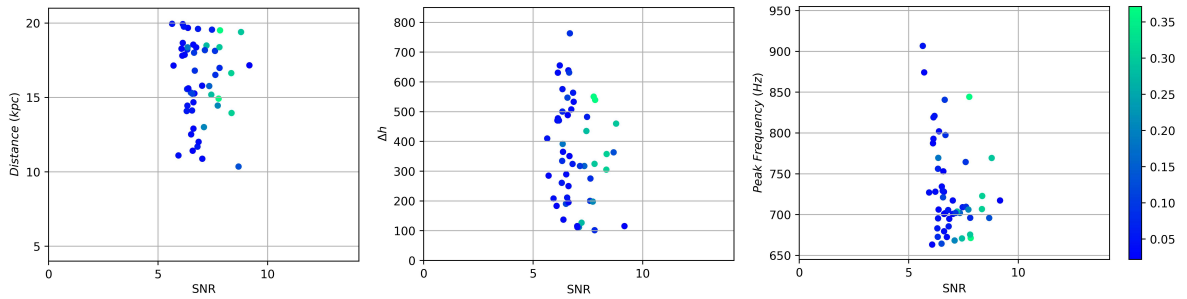


Figure 24: Values of *Distance*, Δh, and f_{peak} as a function of the SNR for the missclassified strains on the dataset of 10k *Time Series*. The color bar represents the score given by the model to each strain.

98%. Figure 25 displays the *confusion matrix* for this test. Only 379 strains of real signals were wrongly classified as noise, which corresponds to 4% of all actual signals.

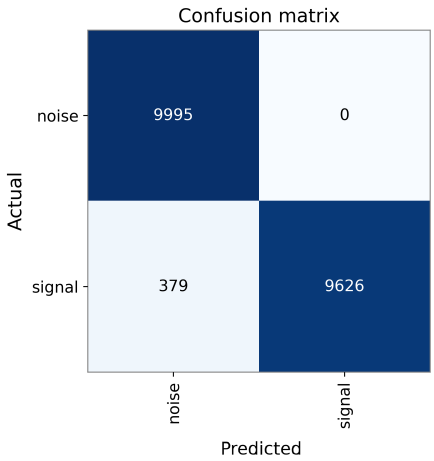


Figure 25: Confusion Matrix for the *classification* test with 10 *epochs* with a dataset of 100k *Time Series*, giving the number of strains predicted as noise or signal versus their actual labels.

The corresponding ROC curve is represented in Figure 26 corresponding to a 0.99 AUC. In addition, Figure 27 displays the distribution of the 379 *False Negatives* obtained in this test. As before, most of the strains have a SNR below 10. We find 8 strains to have a SNR between 10 and 15 and one single strain with 20 SNR, the maximum of this distribution. It is interesting to see that all of these miss-classified strains with a SNR above 10 have a score between 0.30 and 0.50, which shows how the model was not completely certain about the classification predicted. With different values for the threshold, these points would disappear.

6.1.3 TimeSeries vs Spectrograms

In previous work in our group, *classification* tests with DL have already been done with *spectrograms* (time-frequency images) [120], using signals from compact binary coalescences, CCSNe [121], and even

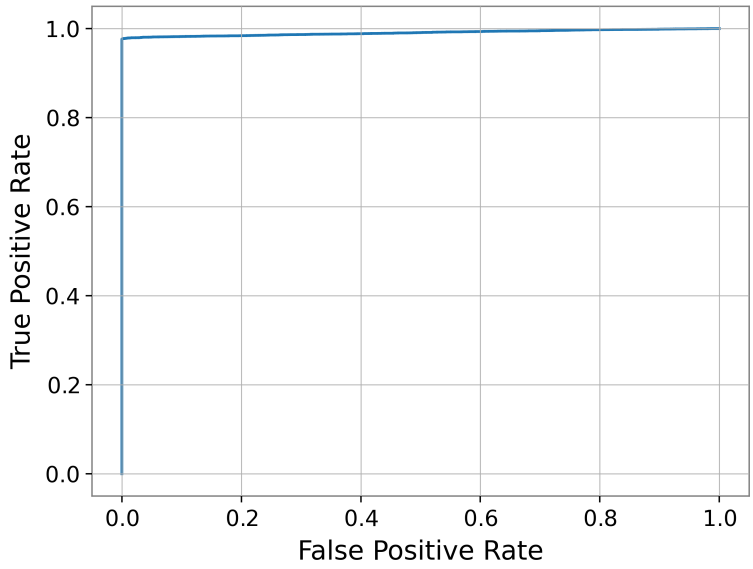


Figure 26: ROC Curve of the *classification* test with 10 *epochs* with a dataset of 100k *Time Series*.

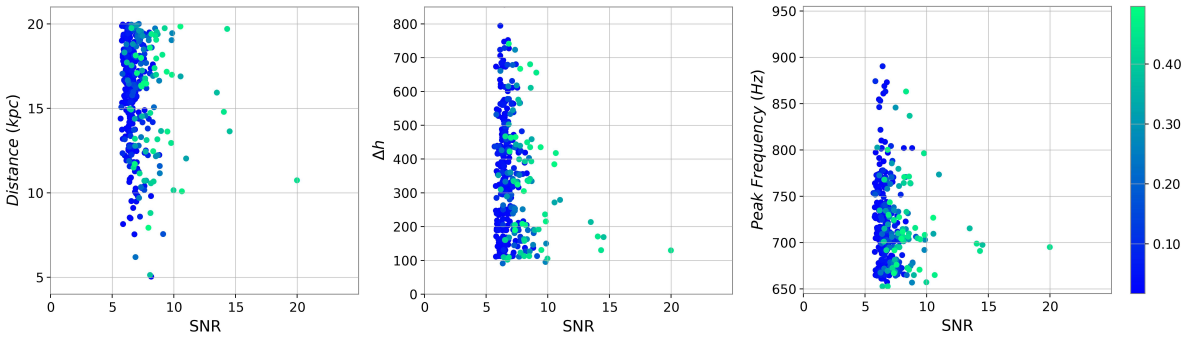


Figure 27: Values of *Distance*, Δh, and f_{peak} as a function of the SNR for the missclassified strains for the dataset of 100k *Time Series*. The color bar represents the score given by the model to each strain.

glitches (i.e. transient sources of noise in the detector of instrumental or environmental origin) [122]. In particular, the CCSNe dataset used by Gabriel Mas in [121] provides a good comparison with the approach using *Time Series* followed in this dissertation, despite the fact that it has different characteristics. In his work, the *distance* was fixed at 20 kpc, using noise from the second observation run (O2) and the *spectrograms* correspond to a time window of 4 seconds.

The *confusion matrix* obtained by Gabriel Mas is displayed in Figure 28. As shown in this figure, 31% of actual signals were classified as background noise and 8% of all background noise were classified as a signal. On the contrary, for the work reported here using *Time Series*, we find that with this dataset only 5% of signals are missclassified as noise and there are no *False Positives*. This improvement in *accuracy* proves that *Time Series* is a good format to use with DL methods, being able to provide results as good as the ones obtained with *spectrograms* or even better.

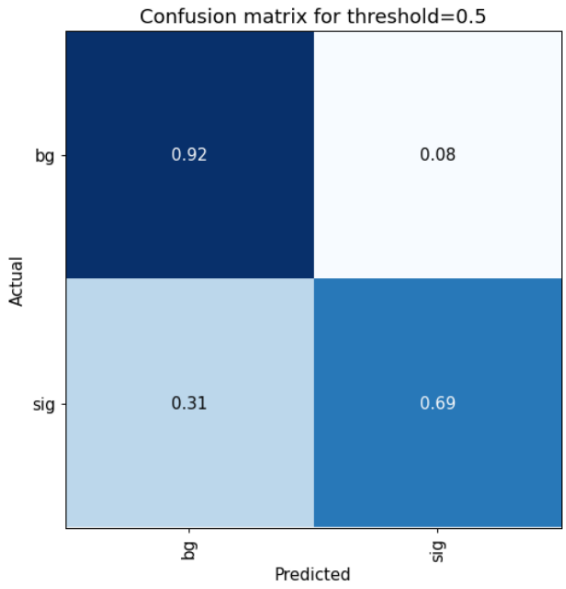


Figure 28: *Confusion Matrix* from the work developed by Gabriel Mas with a dataset of 10k *spectrograms* [121].

6.2 Regression Results

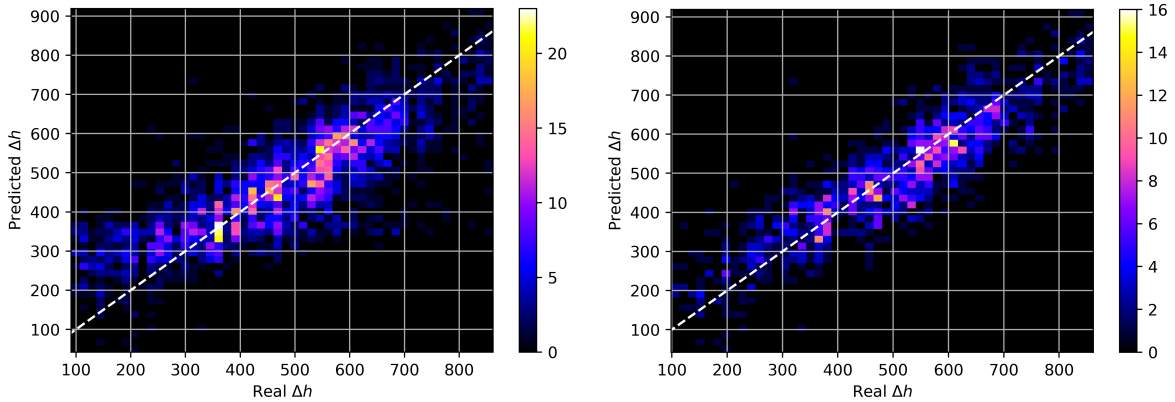
The second major goal of this dissertation is to train a network capable of inferring the physical parameters of our set of CCSNe GWs signals, specifically Δh and f_{peak}.

6.2.1 10k Dataset

As a starting point for the *regression* study, we use a new dataset of 10k *Time Series* with CCSNe signals injected into detector background noise with the same conditions as the ones used for *classification*, presented in Table 1. For this dataset, when performing the *regression tests* the network produces results with a high *Valid Loss*. *Regression* is a more complex study than *classification* and, for this dataset, it was not possible to decrease the value of the *Valid Loss*, even for a higher number of *epochs* when training.

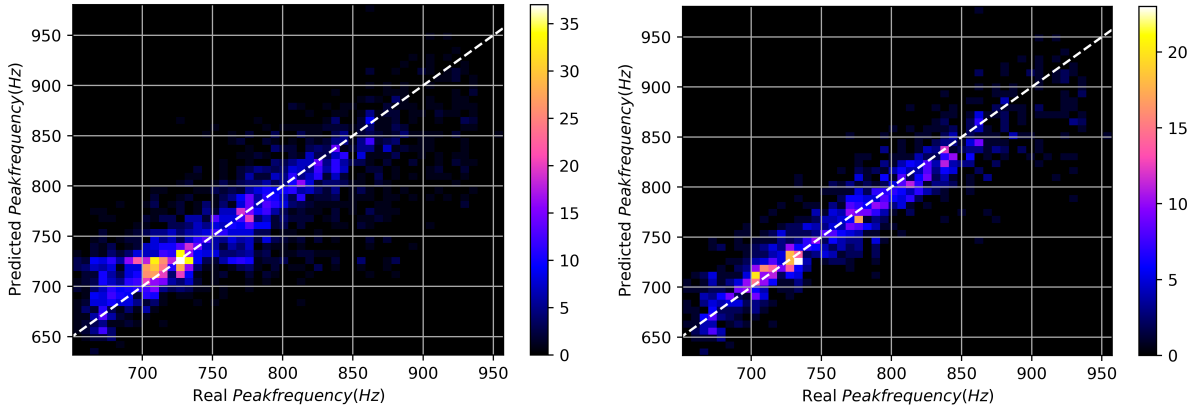
The best network is obtained after training for 50 *epochs* with a 0.30 *Valid Loss*. This training is done with the same maximum LR and *weight decay* of the *classification* tests for the 100k dataset, achieving a *mean absolute error* of 0.40 for this model. Figures 29 and 30 display the behavior of the predictions for the two parameters and for two SNR thresholds in relation to the real values. The distributions follow the white dashed lines which represent the points where the prediction values are equal to the real values ($y = x$). A nice correlation between the two values (real and predicted) is observed in all analyzed distributions.

Comparing the two quantities we perform *regression* on, this model seems to predict the f_{peak} more accurately than the Δh. In Figure 29a a horizontal distribution appears around the value $\Delta h = 350$ and in figure 30a around $f_{\text{peak}} = 725$ Hz. These lines are a consequence of the confusion of the model: when



(a) $SNR \geq 5$ (b) $SNR \geq 25$

Figure 29: Calibration plots with the predicted Δh vs real Δh for values of $SNR \geq 5$ (a) and $SNR \geq 25$ (b), for a dataset of 10k *Time Series*. The white dashed line represents the points where the predicted value is equal to the real value $(y = x)$.



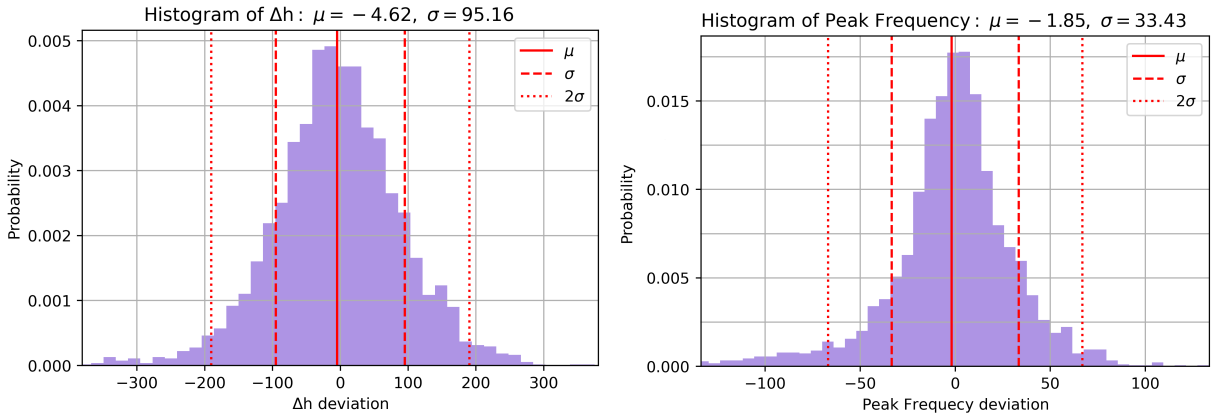
(a) $SNR \geq 5$ (b) $SNR \geq 25$

Figure 30: Calibration plots with the predicted f_{peak} vs real f_{peak} for values of $SNR \geq 5$ (a) and $SNR \geq 25$ (b), for the dataset of 10k *Time Series*. The white dashed line represents the points where the predicted value is equal to the real value $(y = x)$.

the network cannot discern any features in the strains it gives a *mean value* of the real values distribution of these confusing strains. For a minimum SNR value of 5, Figures 29a and 30a, low values of Δh and f_{peak} tend to be slightly overestimated. On the other hand, Figures 29b and 30b show the model's predictions for the strains in the case with a minimum SNR value of 25. In this situation, no horizontal tendency is visible in the plots. This corroborates the trends already observed in the *classification* results: the models, for low SNR, struggle more to separate signal from the noise, which is somehow expected.

The probability distribution of the deviation between real values and predicted values of Δh and f_{peak} is shown in Figure 31. This distribution has a *mean value* of $\mu_{[\Delta h_{\text{real}} - \Delta h_{\text{pred.}}]} = -4.63 \pm 95.16$ for the *bounce strain amplitude* , Δh, and a *mean value* of $\mu_{[f_{\text{peak, real}} - f_{\text{peak, pred.}}]} = -1.85 \pm 33.43$ (Hz) for the *peak*

frequency, f_{peak}. We can clearly see by these values that the *inference* of f_{peak} provides better results than for Δh. Having into account that $\Delta h \in [100, 850]$ and $f_{\text{peak}} \in [650, 950]$ Hz, we can consider reasonable the *mean values* obtained, sustaining the overestimation of the values by the network observed in Figures 29 and 30. On the other hand, the value for the *standard deviation* are too high, specially for Δh. These results were already expected to be under our expectations as value of the *Valid Loss* was still much higher than desired. With a small dataset of only 10k *Time Series* we are limited by the lack of examples and the network does not preform as well as with bigger datasets. Hence, our next step is to repeat this process for a bigger dataset.



(a) (b)

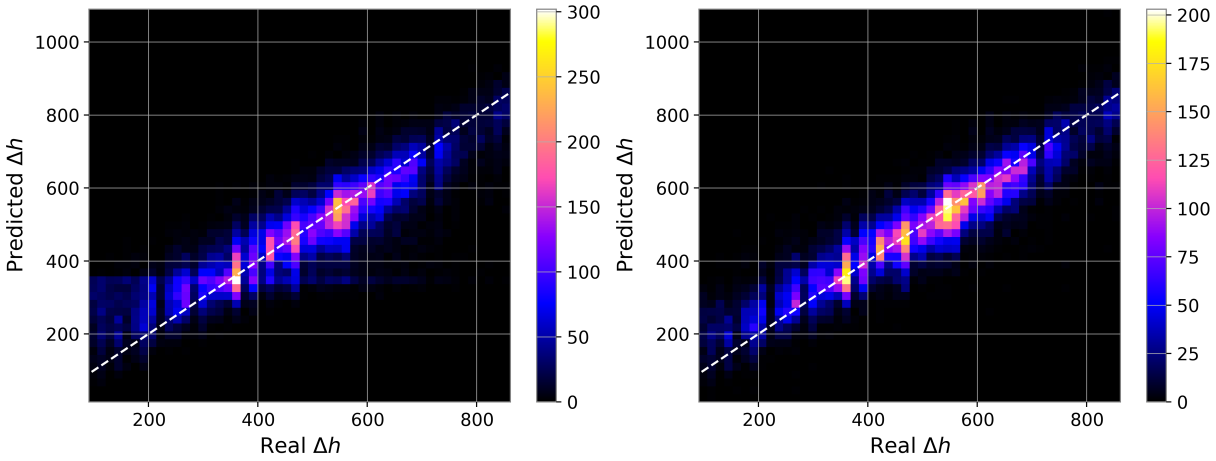
Figure 31: Probability distribution of the deviations of the Δh and f_{peak}, i.e., the difference between real value and predicted value, for the dataset of 10k *Time Series*.

6.2.2 100k Dataset

In order to improve our *regression* results, a bigger dataset was also used, composed of a 100k *Time Series* with the same characteristics as before. In this case, for a training of 50 *epochs* with the previous conditions, the best model is found after 47 *epochs* with 0.230 *Valid Loss*, 23.3% lower than the training with 10k, and 0.335 *mean absolute error*, 16.3% lower than before. Even if it was possible to decrease the *Valid Loss* and the *mean absolute error*, these values are still higher than desired.

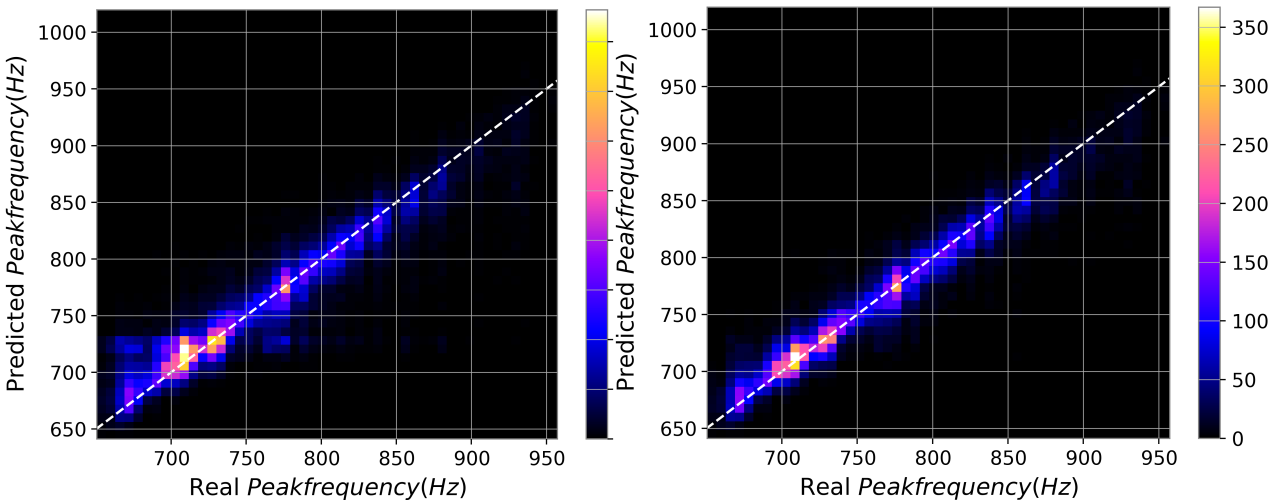
Figures 32 and 33 display the behavior of the predictions in relation to the real values for this new dataset. These distributions are closer to the white dashed lines but the horizontal distributions at the lower left part of the boxes still appear in both plots, Δh and f_{peak}. If only signals with SNR above 15 are selected this horizontal distribution is not visible, as shown in Figure 32b and Figure 33b.

The probability distribution of the deviation between real values and predicted values is shown in Figure 34 for Δh. We obtained $\mu_{[\Delta h_{\text{real}} - \Delta h_{\text{pred.}}]} = -1.141 \pm 81.515$ as the *mean value* of this distribution. When performing these calculations with only a selection of strains, setting a minimum SNR of 15, the *mean value* is $\mu_{[\Delta h_{\text{real}} - \Delta h_{\text{pred.}}]} = 0.625 \pm 70.155$, which shows a slight improvement.



(a) $SNR \geq 5$ (b) $SNR \geq 15$

Figure 32: Calibration plots of predicted Δh vs real Δh for values of $SNR \geq 5$ (a) and $SNR \geq 15$ (b) for a 100k dataset.

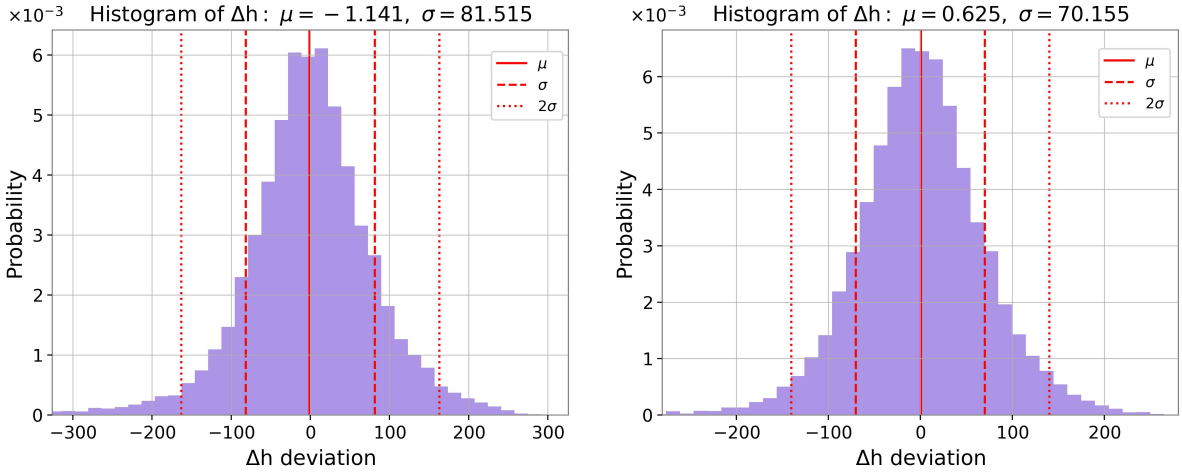


(a) $SNR \geq 5$ (b) $SNR \geq 15$

Figure 33: Calibration plots of predicted f_{peak} vs real f_{peak} for values of $SNR \geq 5$ (a) and $SNR \geq 15$ (b) for a 100k dataset.

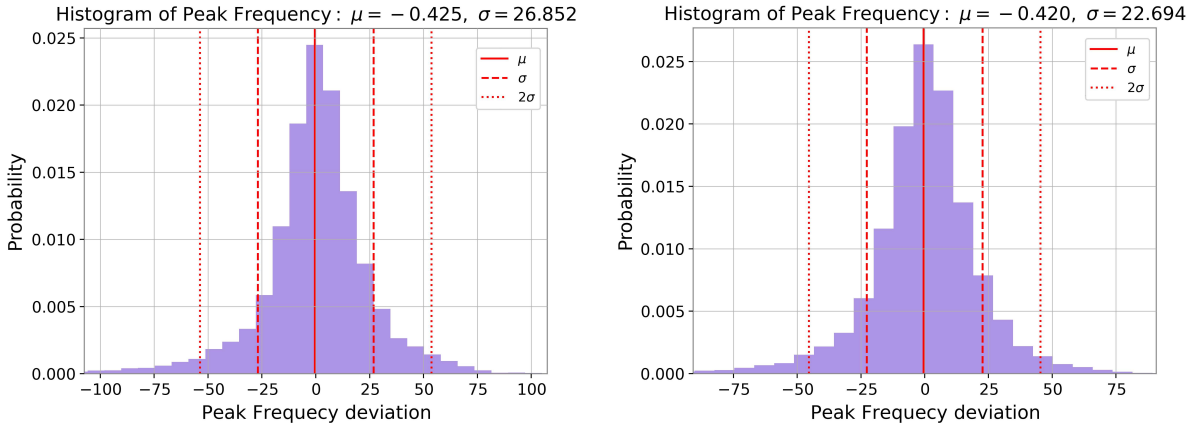
Similarly, the distributions for deviations of the f_{peak}, difference between the real and predicted value, are displayed in Figure 35. We obtained $\mu_{[f_{\text{peak, real}} - f_{\text{peak, pred.}}]} = 0.425 \pm 26.852(\text{Hz})$ for the *mean value* of the differences. The distribution of strains with the minimum SNR of 15 has a *mean value* of $\mu_{[f_{\text{peak, real}} - f_{\text{peak, pred.}}]} = -0.420 \pm 22.694$ (Hz), which also shows improvement.

When looking at these results, we should have into account the current pipeline used in the LVK collaboration for the detection of CCSNe events, coherent Wave Burst (cWB). Szczepanczyk *et al.* tested the *detection efficiency* for distance and SNR in the paper [123]. They used background noise sensitivities



(a) $SNR \geq 5$ (b) $SNR \geq 15$

Figure 34: Probability distribution of the deviations of the Δh for $SNR \geq 5$ (a) and $SNR \geq 15$ (b) for a 100k dataset.



(a) $SNR \geq 5$ (b) $SNR \geq 15$

Figure 35: Probability distribution of the deviations of the f_{peak} for $SNR \geq 5$ (a) and $SNR \geq 15$ (b) for a 100k dataset.

expected for O4 and O5 and signals with similar waveforms to the ones used in this work, i.e. waveforms of GWs signals where the explosion is neutrino-driven. When testing the *detection efficiency* as a function of the distance, they found 50% probability of detecting signals with distances around 10 kpc. In this work, the distance to the source is between 5 and 20 kpc, surpassing the limit mentioned before. In addition, they studied the *detection efficiency* for the SNR and concluded that the 50% probability of detection limit is at 20 SNR. Considering that, in this work, the background noise used is from O3, with lower sensitivity, the results shown in Figure 32 and 33 are expected. In this conditions, requiring parameter inference on such low values of SNR could be too much for the capabilities of the network.

In light of the previous results, a new set of tests were done with datasets composed of signals with the same characteristics as before but with lower cuts in the SNR. By training strains with a minimum

SNR of 10, it is possible to decrease the *Valid Loss* to 0.164, a more acceptable value. The evolution of the network is shown in Figure 36.

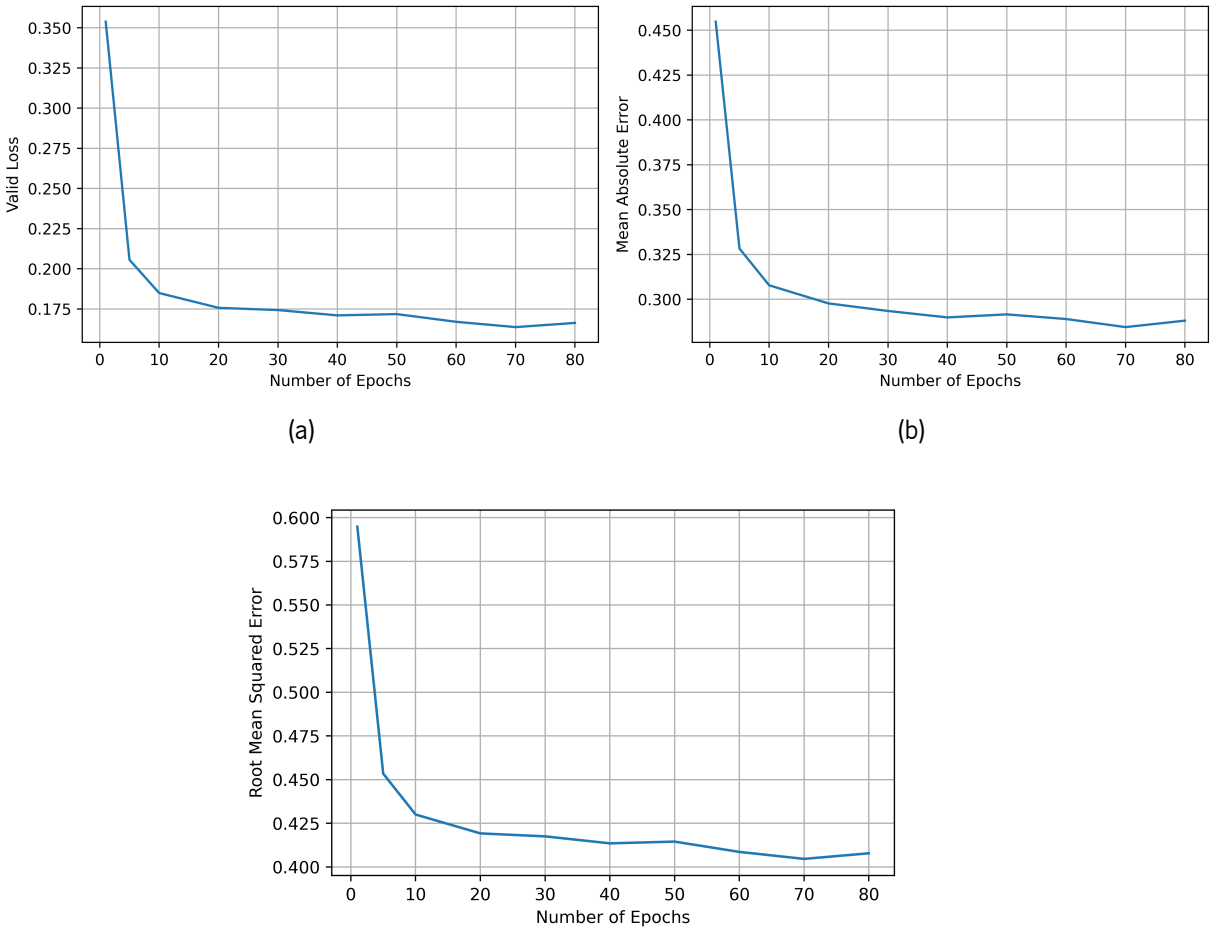


Figure 36: Values of *Valid Loss* (a), *Mean Absolute Error* (b) and *Root Mean Squared Error* (c) obtained when performing different *regression* tests with a different number of *epochs* using a dataset of 100k *Time Series* and setting the minimum SNR as 10.

The best model is found for 70 *epochs* but in the calibration plots the horizontal tendency is still visible. In Figure 37 and 38, the evolution of this horizontal distribution is shown for different minimum values of SNR, for Δh and f_{peak}, respectively. All these tests were done for 70 *epochs*, same LR and *weight decay*, and with a dataset of 100k elements. In particular, we can see in Figure 37 that the horizontal trend completely vanishes for Δh in the case when the training is done with 20 as the minimum value of SNR. As mentioned before, previous studies [123] have showed that with SNR around 20, the current pipeline cWB has a *detection efficiency* of 50%. Having this in mind, our results are completely acceptable and may even be improved by using the expected background noise from O4 or O5. We note that additional white dotted lines are plotted in the four plots of Figure 37 representing the slopes of the distributions, which get closer to the ideal slope of 1. We see an improvement on the slope value, with the enforced SNR

threshold, from 0.832 to 0.923, in the case of Δh. For this distribution, initially we found a *distribution variation* of $\sigma_{\Delta h,5} = 0.342$ which improved to the value $\sigma_{\Delta h,20} = 0.232$ when using 20 as the minimum SNR. Correspondingly, Figure 38 also shows the evolution of the slopes of the distributions with the SNR threshold from 0.857 to 0.951, for the f_{peak}. In addition, for this distribution the *variation* improved from $\sigma_{f_{\text{peak}},5} = 0.306$ Hz to $\sigma_{f_{\text{peak}},20} = 0.174$ Hz. The calculation of the slopes and the *distribution variation* was done with PCA.

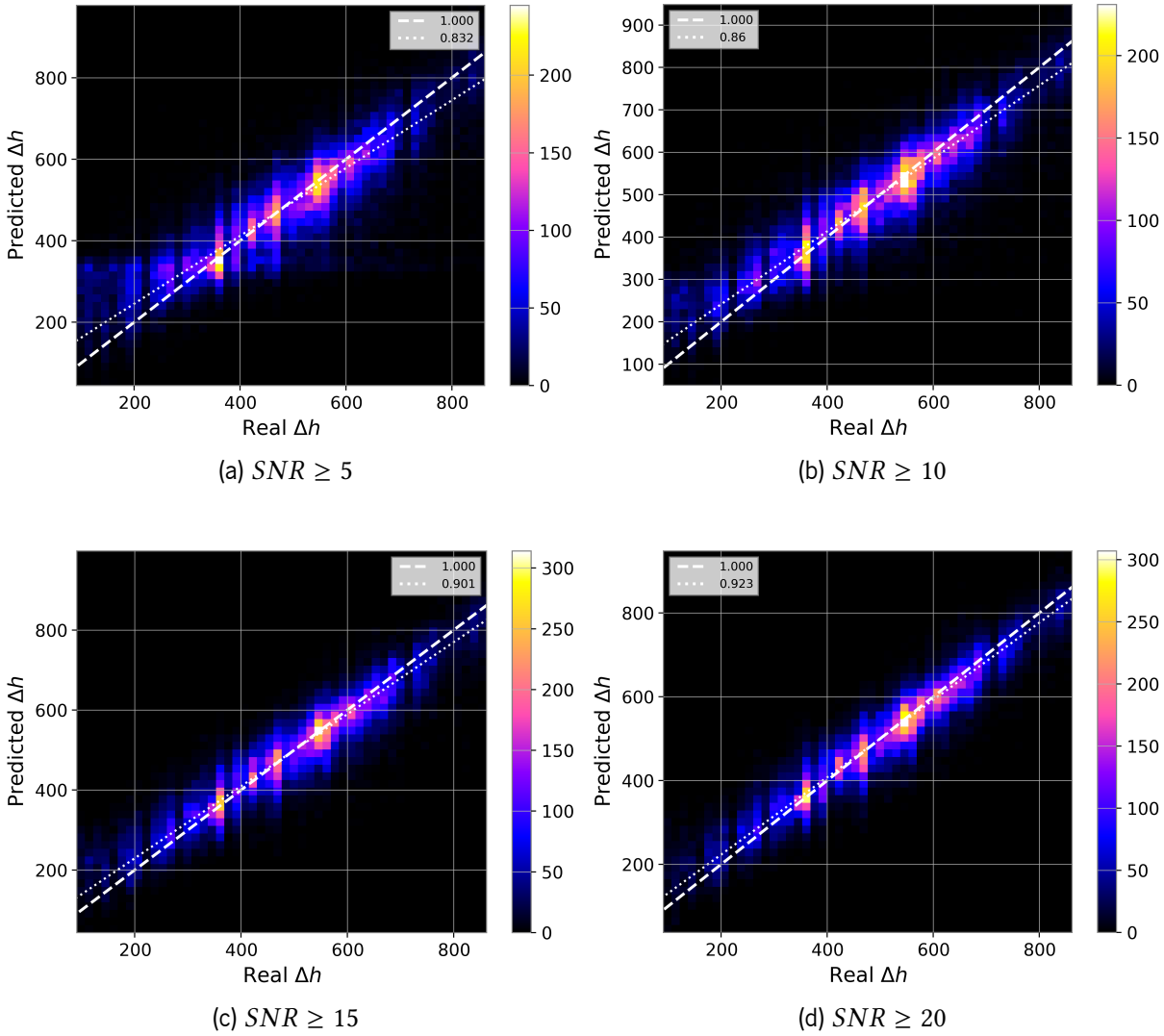
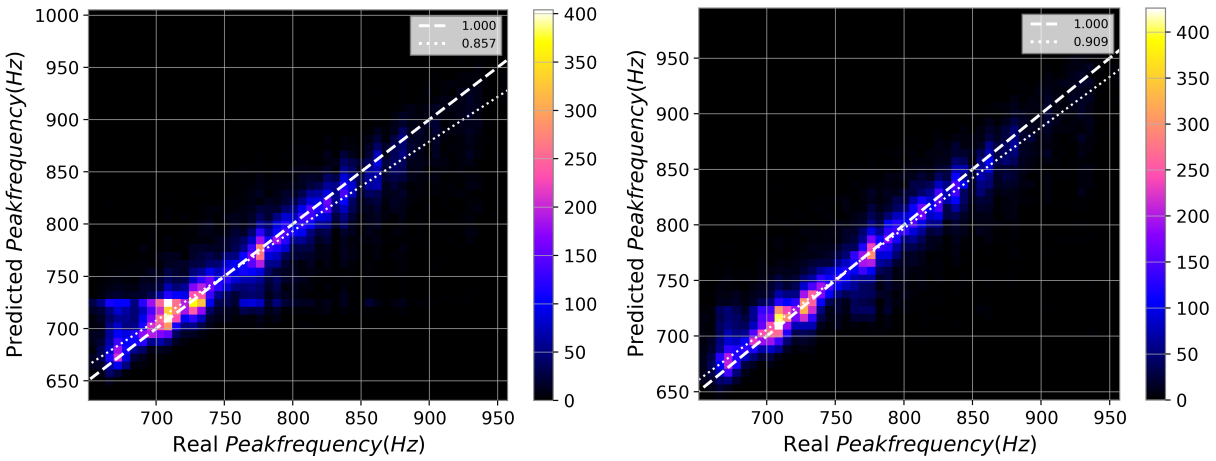
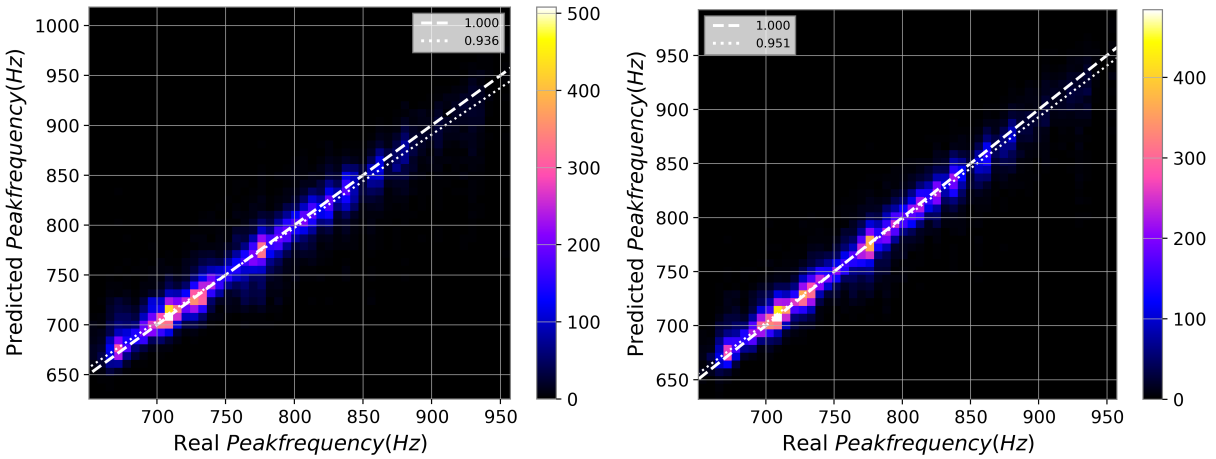


Figure 37: Calibration plots of predicted Δh vs real Δh with $SNR \geq 5$ (a), $SNR \geq 10$ (b), $SNR \geq 15$ (c) and $SNR \geq 20$ (d) for the datasets of 100k *Time Series*.



(a) $SNR \geq 5$

(b) $SNR \geq 10$



(c) $SNR \geq 15$

(d) $SNR \geq 20$

Figure 38: Calibration plots of predicted f_{peak} vs real f_{peak} with $SNR \geq 5$ (a), $SNR \geq 10$ (b), $SNR \geq 15$ (c) and $SNR \geq 20$ (d) for the datasets of 100k *Time Series*.

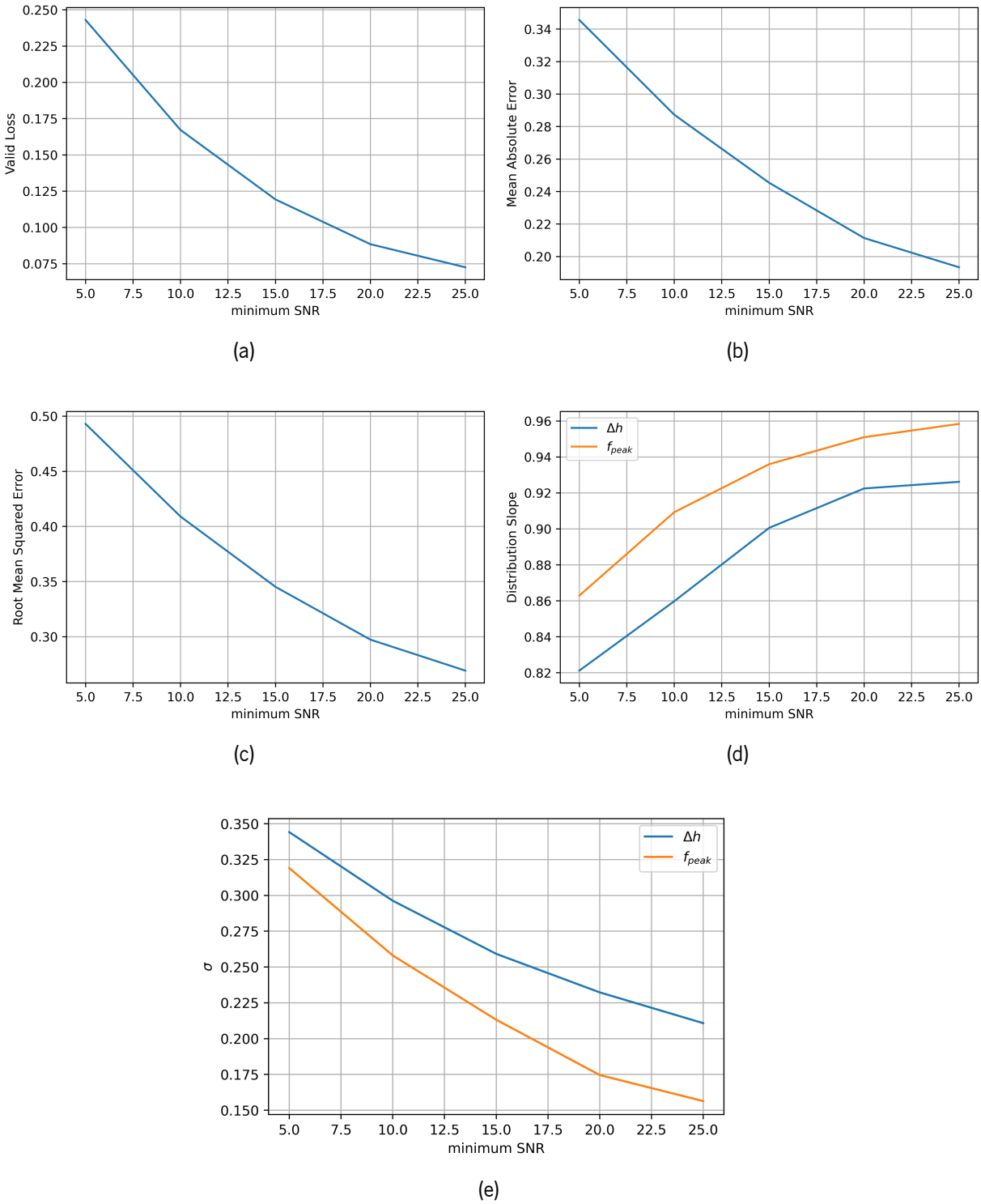
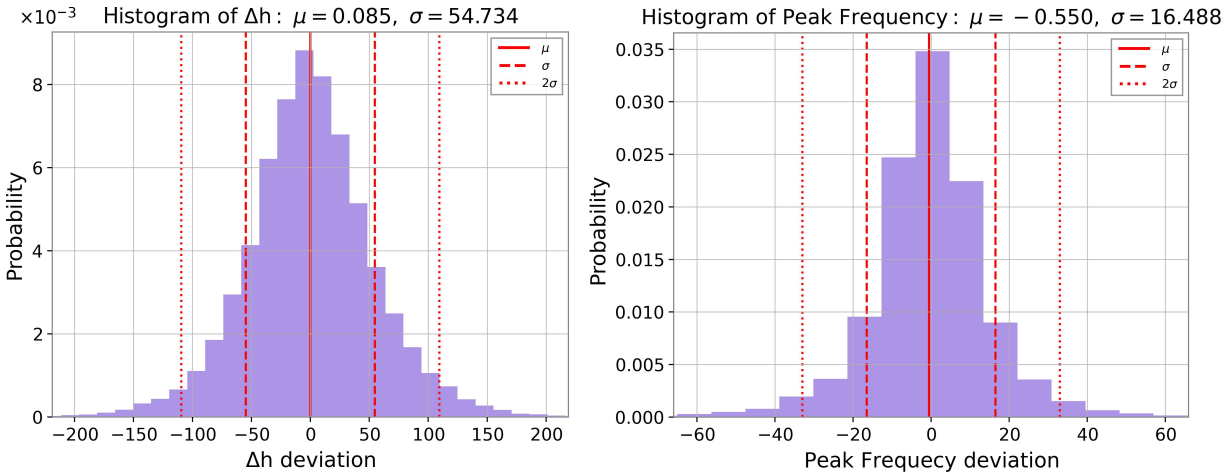


Figure 39: Values of *Valid Loss* (a), *Mean Absolute Error* (b), *Root Mean Squared Error* (c), *Distribution Slope* (d), and *Distribution Variation* (σ) (e) obtained for the different regression tests for different values of the minimum SNR permitted in the datasets of 100k *Time Series*.

With these results, it becomes clear that the SNR of the signals has a significant influence on the training of these networks. In Figure 39 is depicted the evolution of the *Valid Loss*, *Mean Absolute Error*, *Root Mean Squared Error*, *Distribution Slope*, and *Distribution Variation* (σ), for both Δh and f_{peak}. Correspondingly, the probability distribution of the deviations of Δh and f_{peak} for training with the dataset with minimum SNR of 20 can be seen in Figure 40. The two distributions plotted are near zero with much lower values of the *standard deviation*. For Δh the *mean value* is $\mu_{[\Delta h_{\text{real}} - \Delta h_{\text{pred.}}]} = 0.085 \pm 54.734$. For f_{peak}, the *mean value* is $\mu_{[f_{\text{peak, real}} - f_{\text{peak, pred.}}]} = 0.550 \pm 16.488(\text{Hz})$.



(a) (b)

Figure 40: Probability distribution of the deviations of the Δh and f_{peak}, i.e., the difference between real value and predicted value for the dataset of $100k$ with a minimum SNR of 20 with datasets of 100k *Time Series*.

7

Conclusions

The principal objective of this dissertation was to develop ML techniques to perform *classification* and *parameter inference* of rotating CCSNe events using solely the information encoded in the gravitational waveforms they produce. The first detection of a signal from a CCSN is highly anticipated as its gravitational waveform can provide important information about the physical processes occurring during the gravitational collapse of massive stars and about the EOS of the neutron stars that are born through such mechanism. Our goals were achieved by training NNs on numerically generated signals injected into real detector background noise, in the form of *Time Series* (time domain). The results provided in this work offer convincing evidence that DL models are worth considering as alternatives to other approaches (e.g. cWB [124]) and worth implementing in an offline pipeline and, especially, in low latency for CCSNe detection and analysis.

All the tests reported in this work were done with ResCNNs applied to datasets composed of numerically generated 1 second signals from different waveforms at different distances (5-20 kpc) and sky positions, only constraining the inclination at its maximum value, $\pi/2$rad. We started with the *classification* test with a 10k dataset consisting of 50% strains of background noise and 50% strains of injected signals. The network provided a model with an accuracy of 97% with no *false positives*, i.e., the model never classifies strains of noise as actual GWs signals. To verify these results, similar tests were done for a 100k dataset, obtaining a model with 98% accuracy with no *false positives*. These results confirm the efficiency of these networks regarding *classification* and show that, even with a smaller dataset, high values of accuracy can be obtained. With further inspections of the *false negatives*, we could notice a correlation with the SNR of the signals: most of the missed *classifications* occurred for strains with SNRs below 10. This is completely expected, as for lower SNRs it is much more difficult to distinguish signals from the random peaks of the noise, as their amplitudes will be similar.

When comparing our results with previous work of our group with *spectrograms*, we can notice a big improvement in accuracy and performance. The work developed by Gabriel Mas [121] was performed with a *spectrogram* dataset of rotating CCSNe GWs produced with the same catalog but with a fixed distance of 20 kpc, a time window of 4 seconds and background noise from O2. His network miss-classified 8% of noise as signal and 31% of actual signals were classified as noise. We, on the other hand, did not have

any *false positives* and were able to reduce the percentage of actual signals classified as noise to 5%. This proves how *Time Series* are a good format for DL methods, being able to provide results as good as the ones obtained with *spectrograms* or even better.

To perform *parameter inference*, the second major task of the network, we also started by using a 10k dataset, composed only of strains of injected signals. As the resulting model did not converge as well as we wanted to, most likely due to a lack of information in a much too small dataset, all other regression tests were done with a 100k dataset. With a bigger dataset, despite the improvement, the initial results were still not good enough. We can see in the plots reported in Chapter 6, Figure 32 and 33, a tendency for the model to predict the same values in the regression of Δh and *peak frequency*, especially for the lower half of the range of the parameters. With further tests, we also found a correlation between the performance of the network and the SNRs, concluding that the previous horizontal tendency vanishes when the networks are trained with a dataset composed of signals with a lower limit of 20 in the SNR. The studies performed by [123] showed a 50% probability of detecting signals for SNRs above 20 with the current pipeline cWB. We see that our results are in agreement, providing a similar limit for the SNR.

The regression model obtained when introducing a lower limit value for the SNRs of 20, provided for the Δh distribution a mean value of $\mu_{[\Delta h_{\text{real}} - \Delta h_{\text{pred.}}]} = 0.085 \pm 54.734$ for the difference between the real and predicted values. Using PCA, we found this distribution to have a slope of 0.923 and a $\sigma_{\Delta h,20} = 0.232$ distribution variation. Considering now the distribution of the peak frequency, we found a mean value of $\mu_{[f_{\text{peak, real}} - f_{\text{peak, pred}}]} = 0.550 \pm 16.488(\text{Hz})$ for the difference between the real and predicted values. The distribution's slop was 0.923 with $\sigma_{f_{\text{peak}},20} = 0.174$ Hz for the distribution variation.

In light of the improvements planned for current detectors, we expect a significant increase in sensitivity for the next observation runs (O4 and O5) which will lead to a significant rise in the detection rate and in the values of the SNR across the entire spectrum of GWs events. Current GWs search pipelines mostly use matched filtering techniques, which are tailored to compact binary coalescences for which the waveforms are well modelled through analytical and numerical relativity. This approach, however, is not valid for CCSNe events, where the waveforms are short, unmodelled bursts, and the parameter space of the progenitors is much too wide to be covered by numerical simulations. Searches for CCSNe events use coherent multi-detector network time-frequency analysis with minimal assumptions on the morphology of the signals, e.g. cWB [124]. The implementation of DL has already been studied for classification of CCSNe. As an example, [125] applied a Mini-Inception ResNet (also implementing *convolutional layers*) for time-frequency images to improve the detectability of CCSNe GWs signals in cWB. Their work was applied to a dataset of phenomenological CCSNe waveforms, a different type of waveforms from the one used in our work, in background noise from O2. They obtained around 80% efficiency for SNR=16, with a 5% false alarm rate. Our work shows that the use of *Time Series* could possibly decrease the false alarm rate and also increase efficiency, being an interesting complement to those other methods, especially for low-latency detection.

With this in mind, the results reported in this dissertation can still be improved in future work along the following four directions:

- Instead of injecting the numerically simulated signals into background noise from previous observation runs, we could do this research with the background noise expected for the next observation runs or even for future projects like ET [85], and CE [87].
- In this work a limited number of waveforms was used. In the future, the datasets could be more diverse and their parameters extended, in particular incorporating higher distances and varying inclinations. Parameter inference could also be performed for other parameters like the polarization angle.
- We can also test if the models can distinguish glitches from actual CCSNe signals and how well this model would perform in their presence.
- Last but not least, further investigation on DL techniques can be executed by testing different architectures that could possibly provide better results than the ones obtained with ResCNNs.

Bibliography

[1] J. M. Lourenço. *The NOVAthesis LaTeX Template User's Manual*. NOVA University Lisbon. 2021. url: https://github.com/joaomlourenco/novathesis/raw/main/template.pdf (cit. on p. ii).

[2] F. R. Stephenson and D. A. Green. *Historical Supernovae and their Remnants*. https://doi.org/10.1093/acprof:oso/9780198507666.001.0001. OXford Academic, 2002 (cit. on p. 1).

[3] O. Heaviside. "A Gravitational and Electromagnetic Analogy". In: *The Electrician* 31 (1893), pp. 281–282 (cit. on p. 2).

[4] H. Poincaré. "Sur la dynamique de l'électron". In: *Comptes rendus de l'Académie des Sciences* (140 1905), pp. 1504–1508. url: https://www.academie-sciences.fr/pdf/dossiers/Poincare/Poincare_pdf/Poincare_CR1905.pdf (cit. on p. 2).

[5] A. Einstein. "Zur Elektrodynamik bewegter Körper". In: *Annalen der Physik* 322 (10 1905), pp. 891–921. doi: 10.1002/andp.19053221004. url: https://onlinelibrary.wiley.com/doi/10.1002/andp.19053221004 (cit. on p. 2).

[6] A. Einstein. "Die Grundlage der allgemeinen Relativitätstheorie". In: *Annalen der Physik* 354 (7 1916), pp. 769–822. doi: 10.1002/andp.19163540702. url: https://onlinelibrary.wiley.com/doi/10.1002/andp.19163540702 (cit. on pp. 2, 4).

[7] A. Einstein. "Erklärung der Perihelbewegung des Merkur aus der allgemeinen Realtivitätstheorie". In: *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin)* (1915), pp. 831–839. url: https://ui.adsabs.harvard.edu/abs/1915SPAW.....831E/abstract (cit. on p. 2).

[8] F. W. Dyson, A. S. Eddington, and C. Davidson. "A Determination of the Deflection of Light by the Sun's Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a*

Mathematical or Physical Character 220 (1920), pp. 291–333. url: https://www.jstor.org/stable/91137?origin=ads (cit. on p. 2).

[9] R. A. Hulse and J. H. Taylor. "Discovery of a pulsar in a binary system". In: *Astrophysical Journal* 195 (1975), pp. L51–L53. doi: 10.1086/181708 (cit. on p. 2).

[10] J. H. Taylor, L. A. Fowler, and P. M. McCulloch. "Measurements of general relativistic effects in the binary pulsar $PSR1913+16$". In: *Nature* 277 (1979), pp. 437–440. doi: 10.1038/277437a0. url: https://www.nature.com/articles/277437a0 (cit. on p. 2).

[11] M. E. Gertsenshtein and V. I. Pustovoit. "On the Detection of Low-Frequency Gravitational Waves". In: *Soviet Journal of Experimental and Theoretical Physics* 16 (1963), p. 433. url: https://ui.adsabs.harvard.edu/abs/1963JETP...16..433G/abstract (cit. on p. 2).

[12] G. E. Moss, L. R. Miller, and R. L. Forward. "Photon-Noise-Limited Laser Transducer for Gravitational Antenna". In: *Applied Optics* 10 (11 1971), pp. 2495–2498. doi: 10.1364/AO.10.002495. url: https://opg.optica.org/ao/abstract.cfm?uri=ao-10-11-2495 (cit. on p. 3).

[13] F. A. E. Pirani. "Republication of: On the physical significance of the Riemann tensor". In: *General Relativity and Gravitation* 41 (11 2009), pp. 1215–1232. doi: 10.1007/s10714-009-0785-y. url: https://link.springer.com/article/10.1007/s10714-009-0787-9 (cit. on p. 3).

[14] B. Abbott et al. "LIGO: the Laser Interferometer Gravitational-Wave Observatory". In: *Reports on Progress in Physics* 72.7 (2009). doi: 10.1088/0034-4885/72/7/076901. url: https://iopscience.iop.org/article/10.1088/0034-4885/72/7/076901 (cit. on pp. 3, 17).

[15] T. Accadia et al. "Virgo: a laser interferometer to detect gravitational waves". In: *Journal of Instrumentation* 7 (2012). doi: 10.1088/1748-0221/7/03/P03012. url: https://iopscience.iop.org/article/10.1088/1748-0221/7/03/P03012 (cit. on pp. 3, 17).

[16] B. P. Abbott et al. "Observation of Gravitational Waves from a Binary Black Hole Merger". In: *Phys. Rev. Lett.* 116 (6 2016), p. 061102. doi: 10.1103/PhysRevLett.116.061102. url: https://link.aps.org/doi/10.1103/PhysRevLett.116.061102 (cit. on p. 3).

[17] The LIGO Scientific Collaboration, the Virgo Collaboration, and the KAGRA Collaboration. "GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run". In: *arXiv* (2021). doi: 10.48550/arXiv.2111.03606. url: https://arxiv.org/abs/2111.03606# (cit. on p. 3).

[18] T. Akutsu et al. “Overview of KAGRA: Detector design and construction history”. In: *Progress of Theoretical and Experimental Physics* 2021 (5 2021). doi: 10.1093/ptep/ptaa125. url: https://academic.oup.com/ptep/article/2021/5/05A101/5893487 (cit. on pp. 3, 17).

[19] B. P. Abbott et al. “Optically targeted search for gravitational waves emitted by core-collapse supernovae during the first and second observing runs of advanced LIGO and advanced Virgo”. In: *Physical Review D* 101.8, 084002 (2020), p. 084002. doi: 10.1103/PhysRevD.101.084002. arXiv: 1908.03584 [astro-ph.HE] (cit. on p. 3).

[20] N. T. Bishop and L. Rezzolla. “Extraction of Gravitational Waves in Numerical Relativity”. In: *Living Rev Relativ* 19 (2016), p. 2. doi: 10.1007/s41114-016-0001-9. url: https://link.springer.com/article/10.1007/s41114-016-0001-9 (cit. on p. 5).

[21] J. L. Cervantes-Cota et al. “A Brief History of Gravitational Waves”. In: *Universe* 2 (3 2016). doi: 10.3390/universe2030022. url: https://www.mdpi.com/2218-1997/2/3/22 (cit. on p. 4).

[22] S. W. Hawking and W. Israel. *Three Hundred Years of Gravitation*. Cambridge University Press, 1989. isbn: 0521379768 (cit. on p. 4).

[23] N. Christensen. “Stochastic Gravitational Wave Backgrounds”. In: *Rep. Prog. Phys.* 82 (2019). doi: 10.1088/1361-6633/aae6b5. url: https://iopscience.iop.org/article/10.1088/1361-6633/aae6b5 (cit. on p. 4).

[24] M. Maggiore. *Gravitational Waves, volume 1*. Oxford University Press, 2008. isbn: 978-0-19-857074-5 (cit. on p. 5).

[25] A. W. Alsabti and P. Murdin. *Handbook of Supernovae*. https://link.springer.com/referencework/10.1007/978-3-319-20794-0. Springer, 2020 (cit. on p. 10).

[26] R. Minkowski. “Sectra of Supernovae”. In: *Publications of the Astronomical Society of the Pacific* 53 (314 1941-08), p. 224. doi: 10.1086/125315. url: https://ui.adsabs.harvard.edu/abs/1941PASP...53..224M/abstract (cit. on p. 10).

[27] J. C. Wheeler and R. P. Harkness. “Type I supernovae”. In: *Rep. Prog. Phys.* 53 (1467 1990). doi: 10.1088/0034-4885/53/12/001. url: https://iopscience.iop.org/article/10.1088/0034-4885/53/12/001/meta (cit. on p. 10).

[28] R. Barbon, F. Ciatti, and L. Rosino. “Photometric properties of type II supernovae”. In: *Astronomy and Astrophysics* 72 (1979), pp. 287–292. url: https://ui.adsabs.harvard.edu/abs/1979A&A....72..287B (cit. on pp. 10, 11).

[29] E. M. Schlegel. “A new subclass ofType II Supernovae?” In: *Mon. Not. R. astr. Soc.* 244 (1990), pp. 269–271. url: https://ui.adsabs.harvard.edu/abs/1990MNRAS.244..269S (cit. on p. 10).

[30] S. Chandrasekhar. "The Maximum Mass of Ideal White Dwarfs". In: *Astrophysical Journal* 74 (1931), p. 81. doi: 10.1086/146963. url: https://ui.adsabs.harvard.edu/abs/1931ApJ....74...81C (cit. on p. 10).

[31] F. Hoyle and W. A. Fowler. "Nucleosynthesis in Supernovae". In: *Astrophysical Journal* 132 (1960), p. 565. doi: 10.1086/146963. url: https://ui.adsabs.harvard.edu/abs/1960ApJ...132..565H (cit. on p. 10).

[32] R. Abbott et al. "Properties and Astrophysical Implications of the 150 $M_\odot$ Binary Black Hole Merger GW190521". In: *The Astrophysical Journal Letters* 900.1 (2020). doi: 10.3847/2041-8213/aba493. url: https://iopscience.iop.org/article/10.3847/2041-8213/aba493 (cit. on p. 12).

[33] S. E. Woosley and A. Heger. "The Pair-instability Mass Gap for Black Holes". In: *The Astrophysical Journal Letters* 912.2 (2021). doi: 10.3847/2041-8213/abf2c4. url: https://iopscience.iop.org/article/10.3847/2041-8213/abf2c4 (cit. on p. 12).

[34] *CompStar Online Supernovae Equations of State*. https://compose.obspm.fr/. Accessed in 14-04-2023 (cit. on p. 12).

[35] *stellarcollapse.org*. https://stellarcollapse.org/. Accessed in 14-04-2023 (cit. on p. 12).

[36] J. M. Lattimer. "The Nuclear Equation of State and Neutron Star Masses". In: *Annual Review of Nuclear and Particle Science* 62 (2012), pp. 485–515. doi: 10.1146/annurev-nucl-102711-095018. url: https://www.annualreviews.org/doi/abs/10.1146/annurev-nucl-102711-095018 (cit. on p. 13).

[37] S. Richers et al. "Equation of state effects on gravitational waves from rotating core collapse". In: *Physical Review D* 95 (6 2017-03), p. 063019. doi: 10.1103/PhysRevD.95.063019. url: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.95.063019 (cit. on pp. 13, 33–36).

[38] *CoCoNuT Code*. https://www.uv.es/coconut/. Accessed in 5-2-2023 (cit. on pp. 13, 34).

[39] H. Dimmelmeier et al. "Combining spectral and shock-capturing methods: A new numerical approach for 3D relativistic core collapse simulations". In: *Physical Review D* 71 (6 2005-03), p. 064023. doi: 10.1103/PhysRevD.71.064023. url: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.71.064023 (cit. on p. 13).

[40] T. W. Baumgarte and S. L. Shapiro. *Numerical Relativity: Starting from Scratch*. 2021. doi: 10.1017/9781108933445 (cit. on pp. 13, 14).

[41] F. Banyuls et al. “Numerical 3 + 1 General Relativistic Hydrodynamics: A Local Characteristic Approach”. In: *The Astrophysical Journal* 476 (1997), pp. 221–231. doi: 10.1086/303604. url: https://ui.adsabs.harvard.edu/abs/1997ApJ...476..221B/abstract (cit. on pp. 13, 14).

[42] R. Arnowitt, S. Deser, and C. W. Misner. “Republication of: The Dynamics of General Relativity”. In: *Gen Relativ Gravit* 40 (2008), pp. 1997–2027. doi: 10.1007/s10714-008-0661-1. url: https://link.springer.com/article/10.1007/s10714-008-0661-1 (cit. on p. 14).

[43] M. Shibata and Y. Sekiguchi. “Gravitational waves from axisymmetric rotating stellar core collapse to a neutron star in full general relativity”. In: *Physical Review D* 69 (2004), p. 084024. doi: 10.1103/PhysRevD.69.084024. url: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.69.084024 (cit. on pp. 14, 15).

[44] H. Dimmelmeier, J. A. Font, and E. Müller. “Relativistic simulations of rotational core collapse II. Collapse dynamics and gravitational radiation”. In: *Astronomy & Astrophysics* 393 (2002), pp. 523–542. doi: 10.1051/0004-6361:20021053. url: https://www.aanda.org/articles/aa/abs/2002/38/aa2576/aa2576.html (cit. on pp. 14, 15).

[45] P. Cerdá-Durán et al. “CFC+: improved dynamics and gravitational waveforms from relativistic core collapse simulations”. In: *Astronomy and Astrophysics* 3 (439 2005), pp. 1033–1055. doi: 10.1051/0004-6361:20042602. url: https://www.aanda.org/articles/aa/abs/2005/33/aa2602-04/aa2602-04.html (cit. on p. 15).

[46] T. Damour, E. Gourgoulhon, and P. Grandclément. “Circular orbits of corotating binary black holes: Comparison between analytical and numerical results”. In: *Physical Review D* 66 (2002), p. 024007. doi: 10.1103/PhysRevD.66.024007. url: https://journals.aps.org/prd/abstract/10.1103/PhysRevD.66.024007 (cit. on p. 15).

[47] J. Weber. “Observation of the Thermal Fluctuations of a Gravitational-Wave Detector”. In: *Phys. Rev. Lett.* 17 (1966), p. 1228. doi: 10.1103/PhysRevLett.17.1228. url: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.17.1228 (cit. on p. 16).

[48] J. Weber. “Evidence for Discovery of Gravitational Radiation”. In: *Phys. Rev. Lett.* 22 (24 1969), pp. 1320–1324. doi: 10.1103/PhysRevLett.22.1320. url: https://link.aps.org/doi/10.1103/PhysRevLett.22.1320 (cit. on p. 17).

[49] S. T. Kip. *Gravitational Radiation. In Three Hundred Years of Gravitation; Hawking, S., Israel.* https://www.cambridge.org/us/catalogue/catalogue.asp?isbn=0521379768. Cambridge University Press, 1989 (cit. on p. 17).

[50] R. L. Forward. “Wideband laser-interferometer graviational-radiation experiment”. In: *Phys. Rev. D* 17 (2 1978), pp. 379–390. doi: 10.1103/PhysRevD.17.379. url: https://link.aps.org/doi/10.1103/PhysRevD.17.379 (cit. on p. 17).

[51] K. L. Dooley et al. “GEO 600 and the GEO-HF upgrade program: successes and challenges”. In: *Classical and Quantum Gravity* 33.7 (2016). doi: 10.1088/0264-9381/33/7/075009. url: https://iopscience.iop.org/article/10.1088/0264-9381/33/7/075009 (cit. on p. 17).

[52] B. J. Meers. “Recycling in laser-interferometric gravitational-wave detectors”. In: *Phys. Rev. D* 38 (8 1988), pp. 2317–2326. doi: 10.1103/PhysRevD.38.2317. url: https://link.aps.org/doi/10.1103/PhysRevD.38.2317 (cit. on p. 18).

[53] The LIGO Scientific Collaboration. “Advanced LIGO”. In: *Classical and Quantum Gravity* 32.7 (2015). doi: 10.1088/0264-9381/32/7/074001. url: https://iopscience.iop.org/article/10.1088/0264-9381/32/7/074001 (cit. on p. 18).

[54] The VIRGO Collaboration. “Advanced Virgo: a second-generation interferometric gravitational wave detector”. In: *Classical and Quantum Gravity* 32.2 (2014). doi: 10.1088/0264-9381/32/2/024001. url: https://iopscience.iop.org/article/10.1088/0264-9381/32/2/024001 (cit. on p. 18).

[55] A. Buonanno and Y. Chen. “Quantum noise in second generation, signal-recycled laser interferometric gravitational-wave detectors”. In: *Phys. Rev. D* 64 (4 2001), p. 042006. doi: 10.1103/PhysRevD.64.042006. url: https://link.aps.org/doi/10.1103/PhysRevD.64.042006 (cit. on p. 18).

[56] M. Tse et al. “Quantum-Enhanced Advanced LIGO Detectors in the Era of Gravitational-Wave Astronomy”. In: *Phys. Rev. Lett.* 123 (23 2019), p. 231107. doi: 10.1103/PhysRevLett.123.231107. url: https://link.aps.org/doi/10.1103/PhysRevLett.123.231107 (cit. on p. 19).

[57] C. M. Caves. “Quantum-mechanical noise in an interferometer”. In: *Phys. Rev. D* 23 (8 1981), pp. 1693–1708. doi: 10.1103/PhysRevD.23.1693. url: https://link.aps.org/doi/10.1103/PhysRevD.23.1693 (cit. on p. 19).

[58] C. M. Caves. “Quantum-Mechanical Radiation-Pressure Fluctuations in an Interferometer”. In: *Phys. Rev. Lett.* 45 (2 1980), pp. 75–79. doi: 10.1103/PhysRevLett.45.75. url: https://link.aps.org/doi/10.1103/PhysRevLett.45.75 (cit. on p. 19).

[59] Y. Levin. “Internal thermal noise in the LIGO test masses: A direct approach”. In: *Phys. Rev. D* 57 (2 1998), pp. 659–663. doi: 10.1103/PhysRevD.57.659. url: https://link.aps.org/doi/10.1103/PhysRevD.57.659 (cit. on p. 19).

[60] V. Braginsky and S. Vyatchanin. “Thermodynamical fluctuations in optical mirror coatings”. In: *Physics Letters A* 312 (3-4 2003), pp. 244–255. doi: 10.1016/S0375-9601(03)00473-0. url: https://doi.org/10.1016/S0375-9601(03)00473-0 (cit. on p. 19).

[61] W. Yam, S. Gras, and M. Evans. "Multimaterial coatings with reduced thermal noise". In: *Phys. Rev. D* 91 (4 2015), p. 042002. doi: 10.1103/PhysRevD.91.042002. url: https://link.aps.org/doi/10.1103/PhysRevD.91.042002 (cit. on p. 19).

[62] J. Steinlechner. "Development of mirror coatings for gravitational-wave detectors". In: *Phil. Trans. R. Soc. A.* 376 (2120 2018). doi: 10.1098/rsta.2017.0282. url: https://royalsocietypublishing.org/doi/10.1098/rsta.2017.0282 (cit. on p. 19).

[63] G. M. Harry et al. "Titania-doped tantala/silica coatings for gravitational-wave detection". In: *Classical and Quantum Gravity* 24.2 (2006), p. 405. doi: 10.1088/0264-9381/24/2/008. url: https://dx.doi.org/10.1088/0264-9381/24/2/008 (cit. on p. 19).

[64] M. Granata et al. "Amorphous optical coatings of present gravitational-wave interferometers". In: *Classical and Quantum Gravity* 37.9 (2020). doi: 10.1088/1361-6382/ab77e9. url: https://iopscience.iop.org/article/10.1088/1361-6382/ab77e9 (cit. on p. 19).

[65] S. M. Aston et al. "Update on quadruple suspension design for Advanced LIGO". In: *Classical and Quantum Gravity* 29.23 (2012). doi: 10.1088/0264-9381/29/23/235004. url: https://iopscience.iop.org/article/10.1088/0264-9381/29/23/235004 (cit. on p. 19).

[66] S. Wen et al. "Hydraulic external pre-isolator system for LIGO". In: *Classical and Quantum Gravity* 31.23 (2014). doi: 10.1088/0264-9381/31/23/235001. url: https://iopscience.iop.org/article/10.1088/0264-9381/31/23/235001 (cit. on p. 19).

[67] F. Matichard et al. "Seismic isolation of Advanced LIGO: Review of strategy, instrumentation and performance". In: *Classical and Quantum Gravity* 32.18 (2015). doi: 10.1088/0264-9381/32/18/185003. url: https://iopscience.iop.org/article/10.1088/0264-9381/32/18/185003 (cit. on p. 19).

[68] D. Bersanetti et al. "Advanced Virgo: Status of the Detector, Latest Results and Future Prospects". In: *Universe* 7 (9 2021), p. 322. doi: 10.3390/universe7090322. url: https://www.mdpi.com/2218-1997/7/9/322 (cit. on p. 19).

[69] M. G. Beker et al. "Newtonian noise and ambient ground motion for gravitational wave detectors". In: *J. Phys.: Conf. Ser.* 363 (2012). doi: 10.1088/1742-6596/363/1/012004. url: https://iopscience.iop.org/article/10.1088/1742-6596/363/1/012004 (cit. on p. 19).

[70] J. Harms et al. "Observation of a potential future sensitivity limitation from ground motion at LIGO Hanford". In: *Phys. Rev. D* 101 (10 2020), p. 102002. doi: 10.1103/PhysRevD.101.102002. url: https://link.aps.org/doi/10.1103/PhysRevD.101.102002 (cit. on p. 19).

[71] J. Harms. "Terrestrial Gravity Fluctuations". In: *Living Reviews in Relativity* 18 (2015). doi: 10.1007/lrr-2015-. url: https://link.springer.com/article/10.1007/lrr-2015-3 (cit. on p. 19).

[72] J. Harms and K. Venkateswara. "Newtonian-noise cancellation in large-scale interferometric GW detectors using seismic tiltmeters". In: *Classical and Quantum Gravity* 33.23 (2016). doi: 10.1088/0264-9381/33/23/234001. url: https://iopscience.iop.org/article/10.1088/0264-9381/33/23/234001 (cit. on p. 19).

[73] A. Buikema et al. "Sensitivity and performance of the Advanced LIGO detectors in the third observing run". In: *Phys. Rev. D* 102 (6 2020), p. 062003. doi: 10.1103/PhysRevD.102.062003. url: https://link.aps.org/doi/10.1103/PhysRevD.102.062003 (cit. on p. 19).

[74] C. Cahillane and G. Mansell. "Review of the Advanced LIGO Gravitational Wave Observatories Leading to Observing Run Four". In: *Galaxies* 10 (1 2022), p. 36. doi: 10.3390/galaxies10010036. url: https://www.mdpi.com/2075-4434/10/1/36 (cit. on p. 19).

[75] S. J. Cooper et al. "Sensors and actuators for the advanced LIGO A+ upgrade". In: *Rev. Sci. Instrum.* 94 (2023). doi: 10.1063/5.0117605. url: https://aip.scitation.org/doi/full/10.1063/5.0117605 (cit. on p. 19).

[76] F. Acernese et al. "Advanced Virgo Plus: Future Perspectives". In: *Journal of Physics: Conference Series* 2429 (2023). doi: 10.1088/1742-6596/2429/1/012040. url: https://iopscience.iop.org/article/10.1088/1742-6596/2429/1/012040 (cit. on p. 19).

[77] *LIGO, VIRGO and KAGRA Observing Run Plans*. https://observing.docs.ligo.org/plan/. Accessed in 3-03-2023 (cit. on p. 19).

[78] L. McCuller et al. "Frequency-Dependent Squeezing for Advanced LIGO". In: *Phys. Rev. Lett.* 124 (17 2020), p. 171102. doi: 10.1103/PhysRevLett.124.171102. url: https://link.aps.org/doi/10.1103/PhysRevLett.124.171102 (cit. on p. 19).

[79] G. Vajente et al. "Low Mechanical Loss TiO_2 : GeO_2 Coatings for Reduced Thermal Noise in Gravitational Wave Interferometers". In: *Phys. Rev. Lett.* 127 (7 2021), p. 071101. doi: 10.1103/PhysRevLett.127.071101. url: https://link.aps.org/doi/10.1103/PhysRevLett.127.071101 (cit. on p. 20).

[80] *Updated Advanced LIGO sensitivity design curve (T1800044-v5.pdf)*. https://dcc.ligo.org/LIGO-T1800044/public. Accessed in 4-03-2023 (cit. on p. 20).

[81] *Advanced Virgo Technical Design Report (VIR-0128A-12.pdf)*. https://tds.virgo-gw.eu/?content=3&r=9317. Accessed in 4-03-2023 (cit. on p. 20).

[82] M. Saleem, J. Rana, and V. Gayathri et al. “The Science Case for LIGO-India”. In: *Classical and Quantum Gravity* 39.2 (2021). doi: 10.1088/1361-6382/ac3b99. url: https://iopscience.iop.org/article/10.1088/1361-6382/ac3b99 (cit. on p. 21).

[83] J. Baker et al. “The Laser Interferometer Space Antenna: Unveiling the Millihertz Gravitational Wave Sky”. In: *arXiv* (2019). doi: 10.48550/arXiv.1907.06482. url: https://arxiv.org/abs/1907.06482 (cit. on p. 21).

[84] *LISA factsheet*. https://www.esa.int/Science_Exploration/Space_Science/LISA_factsheet. Accessed in 3-03-2023 (cit. on p. 21).

[85] M. Maggiore et al. “Science case for the Einstein telescope”. In: *Journal of Cosmology and Astroparticle Physics* 2020 (2020). doi: 10.1088/1475-7516/2020/03/050. url: https://iopscience.iop.org/article/10.1088/1475-7516/2020/03/050 (cit. on pp. 21, 58).

[86] *ET - Einstein Telescope: Introduction*. https://www.et-gw.eu/index.php. Accessed in 3-03-2023 (cit. on p. 21).

[87] D. Reitze et al. “Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO”. In: *arXiv* (2019). doi: 10.48550/arXiv.1907.04833. url: https://arxiv.org/abs/1907.04833 (cit. on pp. 21, 58).

[88] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016 (cit. on pp. 23, 29).

[89] A. Zhang et al. “Dive into Deep Learning”. In: *arXiv preprint arXiv:2106.11342* (2021) (cit. on p. 24).

[90] L. Alzubaidi and J. Zhang et al. “Review of deep learning: concepts, CNN architectures, challenges, applications, future directions”. In: *Journal of Big Data* 8.53 (2021). doi: 10.1186/s40537-021-00444-8. url: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8#citeas (cit. on p. 26).

[91] A. Khan et al. “A survey of the recent architectures of deep convolutional neural networks”. In: *Artificial Intelligence Review* 53 (2020), pp. 5455–5516. doi: 10.1007/s10462-020-09825-6. url: https://link.springer.com/article/10.1007/s10462-020-09825-6#citeas (cit. on p. 26).

[92] K. He et al. “Deep Residual Learning for Image Recognition”. In: (2015-12). doi: 10.48550/arXiv.1512.03385 (cit. on p. 27).

[93] Y. Gal and Z. Ghahramani. “Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning”. In: (2016-10). doi: 10.48550/arXiv.1506.02142 (cit. on p. 29).

[94] L. N. Smith. “Cyclical Learning Rates for Training Neural Networks”. In: (2017-04). doi: 10.48550/arXiv.1506.01186 (cit. on pp. 30, 31, 39).

[95] L. N. Smith and N. Topin. “Super-Convergence: Very Fast Training of Neural Networks Using Large Learning Rates”. In: (2018-05). doi: 10.48550/arXiv.1708.07120 (cit. on pp. 30, 32).

[96] M.-A. Bizouard et al. “Inference of protoneutron star properties from gravitational-wave data in core-collapse supernovae”. In: *Physical Review D* 103.6 (2021), p. 063006. doi: 10.1103/PhysRevD.103.063006. url: https://ui.adsabs.harvard.edu/abs/2021PhRvD.103f3006B/abstract (cit. on p. 33).

[97] S. Richers et al. *Equation of state effects on gravitational waves from rotating core collapse*. http://dx.doi.org/10.5281/zenodo.201145. Accessed in 8-10-2022 (cit. on p. 33).

[98] J. M. Lattimer and F. D. Swesty. “A generalized equation of state for hot, dense matter”. In: *Nuclear Physics A* 535 (2 1991), pp. 331–376. doi: 10.1016/0375-9474(91)90452-C. url: https://www.sciencedirect.com/science/article/abs/pii/037594749190452C (cit. on p. 33).

[99] H. Shen et al. “Relativistic equation of state of nuclear matter for supernova and neutron star”. In: *Nuclear Physics A* 637 (3 1998), pp. 435–450. doi: 10.1016/S0375-9474(98)00236-X. url: https://www.sciencedirect.com/science/article/pii/S037594749800236X (cit. on p. 33).

[100] H. Shen et al. “Relativistic Equation of State of Nuclear Matter for Supernova Explosion”. In: *Progress of Theoretical Physics* 100 (5 1998), pp. 1013–1031. doi: 10.1143/PTP.100.1013. url: https://academic.oup.com/ptp/article/100/5/1013/1867001 (cit. on p. 33).

[101] H. Shen et al. “Relativistic Equation of State for Core-Collapse Supernova Simulations”. In: *The Astrophysical Journal Supplement Series* 197.2 (2011). doi: 10.1088/0067-0049/197/2/20. url: https://iopscience.iop.org/article/10.1088/0067-0049/197/2/20 (cit. on p. 33).

[102] M. Hempel and J. Schaffner-Bielich. “A statistical model for a complete supernova equation of state”. In: *Nuclear Physics A* 837 (3-4 2010), pp. 210–254. doi: 10.1016/j.nuclphysa.2010.02.010. url: https://www.sciencedirect.com/science/article/pii/S0375947410003325 (cit. on p. 33).

[103] M. Hempel et al. “New Equations of State in Simulations of Core-Collapse Supernovae”. In: *The Astrophysical Journal* 748.1 (2012). doi: 10.1088/0004-637X/748/1/70. url: https://iopscience.iop.org/article/10.1088/0004-637X/748/1/70 (cit. on p. 33).

[104] A. W. Steiner, M. Hempel, and T. Fischer. “Core-collapse supernova equations of state based on neutron star observations”. In: *The Astrophysical Journal* 774.1 (2013). doi: 10.1088/0004-637X/774/1/17. url: https://iopscience.iop.org/article/10.1088/0004-637X/774/1/17 (cit. on p. 33).

[105] S. Banik et al. “New Hyperon Equations of State for Supernovae and Neutron Stars in Density-dependent Hadron Field Theory”. In: *The Astrophysical Journal Supplement Series* 214.2 (2014). doi: 10.1088/0067-0049/214/2/22. url: https://iopscience.iop.org/article/10.1088/0067-0049/214/2/22 (cit. on p. 33).

[106] G. Shen, C. J. Horowitz, and S. Teige. “New equation of state for astrophysical simulations”. In: *Phys. Rev. C* 83 (3 2011), p. 035802. doi: 10.1103/PhysRevC.83.035802. url: https://link.aps.org/doi/10.1103/PhysRevC.83.035802 (cit. on p. 33).

[107] G. Shen, C. J. Horowitz, and E. O’Connor. “Second relativistic mean field and virial equation of state for astrophysical simulations”. In: *Phys. Rev. C* 83 (6 2011), p. 065808. doi: 10.1103/PhysRevC.83.065808. url: https://link.aps.org/doi/10.1103/PhysRevC.83.065808 (cit. on p. 33).

[108] E. O’Connor. “An open-source neutrino hydrodynamics code for core-collapse supernovae”. In: *The Astrophysical Journal Supplement Series* 219 (24 2015-08). doi: 10.1088/0067-0049/219/2/24. url: https://https://iopscience.iop.org/article/10.1088/0067-0049/219/2/24 (cit. on p. 34).

[109] M. Liebendörfer. “A Simple Parameterization of the Consequences of Deleptonization for Simulations of Stellar Core Collapse”. In: *The Astrophysical Journal* 633 (2 2005-11), p. 1042. doi: 10.1086/466517. url: https://iopscience.iop.org/article/10.1086/466517 (cit. on p. 34).

[110] *python*. https://www.python.org. Accessed in 14-04-2023 (cit. on p. 37).

[111] *SciPy*. https://scipy.org. Accessed in 14-04-2023 (cit. on p. 37).

[112] *PyCBC*. https://pycbc.org. Accessed in 14-04-2023 (cit. on p. 37).

[113] *GWpy*. https://gwpy.github.io. Accessed in 14-04-2023 (cit. on p. 37).

[114] *fast.ai*. https://www.fast.ai. Accessed in 14-04-2023 (cit. on p. 37).

[115] *tsai: State-of-the-art Deep Learning library for Time Series and Sequences*. https://timeseriesai.github.io/tsai/. Accessed in 14-04-2023 (cit. on p. 37).

[116] X. Zou et al. “Integration of residual network and convolutional neural network along with various activation functions and global pooling for time series classification”. In: *Neurocomputing* 367 (2019-11). doi: 10.1016/j.neucom.2019.08.023. url: https://doi.org/10.1016/j.neucom.2019.08.023 (cit. on p. 38).

[117] L. N. Smith. “A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay”. In: (2018-04). doi: 10.48550/arXiv.1803.09820 (cit. on p. 39).

[118] *scikit-learn: Machine Learning in Python*. https://scikit-learn.org/stable/. Accessed in 14-04-2023 (cit. on p. 40).

[119] Universitat de València. *ARTEMISA*. https://artemisa.ific.uv.es/web/. Accessed in 16-10-2022 (cit. on p. 40).

[120] J. D. Álvares and J. A. Font et al. "Exploring gravitational-wave detection and parameter inference using deep learning methods". In: *Classical and Quantum Gravity* 38.15 (2021). doi: 10.1088/1361-6382/ac0455. url: https://iopscience.iop.org/article/10.1088/1361-6382/ac0455 (cit. on p. 44).

[121] G. E. Mas. "Trabajo de Fin de Grado en Física: Inferencia de señales gravitatorias de supernovas mediante técnicas de Deep Learning". In: (2021) (cit. on pp. 44–46, 56).

[122] T. S. Fernandes and S. J. Vieira et al. "Convolutional Neural Networks for the classification of glitches in gravitational-wave data streams". In: *arXiv* (2023). doi: 10.48550/arXiv.2303.13917. url: https://arxiv.org/abs/2303.13917 (cit. on p. 45).

[123] M. J. Szczepa□czyk et al. "Detecting and reconstructing gravitational waves from the next galactic core-collapse supernova in the advanced detector era". In: *Physical Review D* 104.10, 102002 (2021), p. 102002. doi: 10.1103/PhysRevD.104.102002. arXiv: 2104.06462 [astro-ph.HE] (cit. on pp. 49, 51, 57).

[124] M. Drago et al. "coherent WaveBurst, a pipeline for unmodeled gravitational-wave data analysis". In: *SoftwareX* 14 (2021-06), p. 100678. doi: 10.1016/j.softx.2021.100678. url: https://www.softxjournal.com/article/S2352-7110(21)00023-6/fulltext (cit. on pp. 56, 57).

[125] M. López et al. "Deep learning for core-collapse supernova detection". In: *Physical Review D* 103.6, 063011 (2021), p. 063011. doi: 10.1103/PhysRevD.103.063011 (cit. on p. 57).