



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÕES



sid.inpe.br/mtc-m21d/2021/09.06.19.25-TDI

ADAPTIVE MULTIREOLUTION FOR WAVELET ANALYSIS

Ana Luiza Bastos Barbosa Guimarães da Silva

Master's Dissertation of the
Graduate Course in Astrophysics,
guided by Drs. Odylio Denys de
Aguiar, and Riccardo Sturani,
approved in July 14, 2021.

URL of the original document:

<<http://urlib.net/8JMKD3MGP3W34T/45CUBSS>>

INPE
São José dos Campos
2021

PUBLISHED BY:

Instituto Nacional de Pesquisas Espaciais - INPE
Coordenação de Ensino, Pesquisa e Extensão (COEPE)
Divisão de Biblioteca (DIBIB)
CEP 12.227-010
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/7348
E-mail: pubtc@inpe.br

**BOARD OF PUBLISHING AND PRESERVATION OF INPE
INTELLECTUAL PRODUCTION - CEPPII (PORTARIA Nº
176/2018/SEI-INPE):****Chairperson:**

Dra. Marley Cavalcante de Lima Moscati - Coordenação-Geral de Ciências da Terra
(CGCT)

Members:

Dra. Ieda Del Arco Sanches - Conselho de Pós-Graduação (CPG)
Dr. Evandro Marconi Rocco - Coordenação-Geral de Engenharia, Tecnologia e
Ciência Espaciais (CGCE)
Dr. Rafael Duarte Coelho dos Santos - Coordenação-Geral de Infraestrutura e
Pesquisas Aplicadas (CGIP)
Simone Angélica Del Ducca Barbedo - Divisão de Biblioteca (DIBIB)

DIGITAL LIBRARY:

Dr. Gerald Jean Francis Banon
Clayton Martins Pereira - Divisão de Biblioteca (DIBIB)

DOCUMENT REVIEW:

Simone Angélica Del Ducca Barbedo - Divisão de Biblioteca (DIBIB)
André Luis Dias Fernandes - Divisão de Biblioteca (DIBIB)

ELECTRONIC EDITING:

Ivone Martins - Divisão de Biblioteca (DIBIB)
André Luis Dias Fernandes - Divisão de Biblioteca (DIBIB)



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÕES



sid.inpe.br/mtc-m21d/2021/09.06.19.25-TDI

ADAPTIVE MULTIREOLUTION FOR WAVELET ANALYSIS

Ana Luiza Bastos Barbosa Guimarães da Silva

Master's Dissertation of the
Graduate Course in Astrophysics,
guided by Drs. Odylio Denys de
Aguiar, and Riccardo Sturani,
approved in July 14, 2021.

URL of the original document:

<<http://urlib.net/8JMKD3MGP3W34T/45CUBSS>>

INPE
São José dos Campos
2021

Cataloging in Publication Data

Silva, Ana Luiza Bastos Barbosa Guimarães da.
Si38a Adaptive Multiresolution for Wavelet Analysis / Ana Luiza
Bastos Barbosa Guimarães da Silva. – São José dos Campos :
INPE, 2021.
xxiv + 85 p. ; (sid.inpe.br/mtc-m21d/2021/09.06.19.25-TDI)

Dissertation (Master in Astrophysics) – Instituto Nacional de
Pesquisas Espaciais, São José dos Campos, 2021.

Guiding : Drs. Odylio Denys de Aguiar, and Riccardo Sturani.

1. Gravitational wave. 2. Data analysis. 3. Signal. 4. Time-
frequency representations. 5. Wavelets. I.Title.

CDU 551.511.31:52



Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](https://creativecommons.org/licenses/by-nc/3.0/).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/).



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA
E INOVAÇÕES



INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS
Serviço de Pós-Graduação - SEPGR

DEFESA FINAL DE DISSERTAÇÃO DE ANA LUIZA BASTOS BARBOSA GUIMARÃES DA SILVA
BANCA Nº 187/2021, REG 208844/2019

No dia 14 de julho de 2021, as 13h00min, por teleconferência, o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi APROVADO(A) pela Banca Examinadora, por unanimidade, em cumprimento ao requisito exigido para obtenção do Título de Mestra em Astrofísica. O trabalho precisa da incorporação das correções sugeridas pela Banca Examinadora e revisão final pelo(s) orientador(es).

Título: “ADAPTIVE MULTIREOLUTION FOR WAVELET ANALYSIS”

Membros da banca:

Dr. Francisco José Jablonski - Presidente - INPE
Dr. Odylio Denys de Aguiar - Orientador - INPE
Dr. Riccardo Sturani - Orientador – UFRN
Dr. Rafael da Costa Nunes – Membro Interno – INPE
Dra. Iara Tosta e Melo – Membro Externo - INFN-LNS



Documento assinado eletronicamente por **Francisco Jose Jablonksi, Pesquisador**, em 15/07/2021, às 08:47 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Odylio Denys de Aguiar, Pesquisador**, em 15/07/2021, às 17:00 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rafael da Costa nunes (E), Usuário Externo**, em 21/07/2021, às 11:01 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Iara Tosta e Melo (E), Usuário Externo**, em 21/07/2021, às 11:15 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Riccardo Sturani (E), Usuário Externo**, em 29/07/2021, às 11:25 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site <http://sei.mctic.gov.br/verifica.html>, informando o código verificador **7879449** e o código CRC **A319CFE5**.

Referência: Processo nº 01340.004555/2021-75

SEI nº 7879449

“Look up at the stars and not down at your feet. Try to make sense of what you see, and wonder about what makes the universe exist. Be curious.”

STEPHEN HAWKING

*To my parents **Elainy** and **Maurício**, to my "vovó"
Terezinha and In Memoriam of my "vovô" **Airton***

ACKNOWLEDGEMENTS

First, I would like to thank both my advisors, professor Odylio Aguiar and professor Riccardo Sturani for all the help, dedication, patience, guidance and understanding. Without whom, I would not be here presenting this work.

To all INPE employees, students and staff, especially to my classmates Denimara and Juliédson for all the hours of studies spend together, laughs, dancing lessons, ice creams, to all the good and bad moments shared together inside and outside the Institute and for sharing the same love for astrophysics. To "Vitol" and Larissa, for being great friends to me in times of need and also for the amazing moments I got to spend with each of them. I have grown professionally and personally with each of them in my life and I couldn't be more grateful.

To my science siblings: Vanessa, who's always with me in good and bad times, who was and still is the best work partner I've ever had and to Heron, for presenting me the beauty in physics and for the help that lead me to choose the path I'm taking in science today. To all the support, friendship and love both of them have given me through all these years.

To my parents, Elaine and Maurício, for supporting me in every moment, encouraging me to pursue my dreams, for all the advises and care. They are my absolute foundation in life.

In memoriam of my grandfather Airton, who left this world too soon, but believed in me right until the end. To my grandmother Terezinha, for all the great stories, snacks and moments we still share together. Every work I'll do in life will be to make them proud of me.

To CAPES, for the funding and investment in science students who like me, are pursuing their dream.

ABSTRACT

Since the first gravitational wave detection in 2015, a new field in astrophysics was opened. A long way was taken for that detection to be made, a theoretical as well as an experimental effort. But most importantly the analysis of the data generated by the gravitational wave detectors (interferometers, such as LIGO and Virgo) is crucial. The data can be described in time-domain and frequency-domain, but none of them offers a good representation of time and frequency at the same time, because of the Heisenberg-Gabor (HB) Principle. Another attempt is to use Wavelets, that is, to use time series as an input for the Wavelet Transform that will represent the data in time-frequency (wavelet domain) with "tiles", also known as Heisenberg Boxes with sides given by time and frequency resolutions, respectively. The wavelet method minimizes the area of the tile defined by the HB Principle. Therefore a code in Python is developed to use any time series as input, to process the data (signal+noise) with windowing, whitening, applying filters (an optional step) and finally performing the Discrete Wavelet Transform (DWT). The DWT output is a list of coefficients, each of which refers to a time bin and a frequency bin, the coefficients can be filtered (another optional step) when using a fixed quantile threshold. Normally, the DWT stops when it reaches the maximum level, but in this work a selection criterion for the decompositions is provided using Shannon's Entropy Criterion that chooses which time and frequency resolutions will be used to represent the input, given their product is constant. The lower the entropy, more the coefficient's power is concentrated in a lower number of tiles. The multiresolution representation adjusts itself to make the best representation of every time series used as input. To test the developed algorithm, the code was applied to 10 gravitational wave events detected and published by the LIGO Scientific Collaboration (LSC), a time-frequency signal was simulated from the best fit for the input and was superimposed to the wavelet representations for comparison. A histogram of the decomposition's coefficients were made for each event, to qualify the goodness of the representations. As a result of that, 6 events steps out as the ones that allows a better discrimination between background and foreground. The most effective signal identifications of our method happens for signals with large frequency variability, as for signals generated by late inspiral and merger of compact binary coalescence.

Keywords: Gravitational Wave. Data Analysis. Signal. Time-frequency representations. Wavelets. Shannon's Entropy Criterion.

MULTIRESOLUÇÃO ADAPTATIVA PARA ANÁLISE WAVELET

RESUMO

Desde a primeira detecção de onda gravitacional em 2015, um novo ramo na astrofísica se abriu. Um longo caminho foi percorrido para que a detecção fosse realizada, tanto um esforço teórico quanto experimental. Mas, mais importante a análise de dados gerados pelos detectores de ondas gravitacionais (interferômetros, como o LIGO e o Virgo) é essencial. Os dados são descritos no domínio do tempo e no domínio da frequência, mas nenhum deles oferece uma boa representação de tempo e frequência ao mesmo tempo, devido ao Princípio de Heisenberg-Gabor (HB). Outra tentativa é utilizar Wavelets, isto é, usar séries temporais como input para a Transformada Wavelet que irá representar os dados em tempo-frequência (domínio wavelet) com "azulejos", também conhecidos como "Caixas de Heisenberg" com lados dados pela resolução no tempo e resolução em frequência, respectivamente. O método wavelet minimiza a área desse azulejo definido pelo Princípio de HB. Portanto um código em Python é desenvolvido para usar qualquer série temporal como input, para processar os dados (sinal+ruído) com janelamento, branqueamento, aplicando filtros (passo opcional) e finalmente executando a Transformada Wavelet Discreta (TWD). O output da TWD é uma lista de coeficientes, cada um deles se refere a um bin de tempo e um bin de frequência, os coeficientes são filtrados (outro passo opcional) quando utiliza-se um limiar fixo de quantil. Normalmente, a TWD para quando alcança o nível máximo, mas neste trabalho um critério de seleção é dado pelo Critério de Entropia de Shannon que seleciona quais resoluções de tempo e frequência serão utilizados na representação do input, dado o produto deles é constante. Quanto menor a entropia, mais a potência do coeficiente está concentrado em um menor número de azulejos. A representação do sinal com multiresolução se adapta para gerar a melhor representação de todas séries temporais usadas como input. Para testar o algoritmo desenvolvido, o código foi aplicado à 10 eventos de ondas gravitacionais detectadas e publicadas pela Colaboração Científica do LIGO (CCL), um sinal em tempo-frequência foi simulado pelo melhor fit para o input e superposto nas representações wavelet para comparação. um histograma dos coeficientes da decomposição foram feitos para cada evento, para quantificar e estabelecer padrões para representações boa e mediana. Como resultado disso, 6 eventos se destacam como os que permitem uma discriminação maior entre background e foreground. As identificações de sinais mais eficazes de nosso método acontecem para sinais com grande variabilidade de frequência, como para sinais gerados por inspiral tardio e fusão de binária compacta coalescente.

Palavras-chave: Ondas Gravitacionais. Análise de dados. Sinal. Representações em tempo-frequência. Wavelets. Critério de Entropia de Shannon.

LIST OF FIGURES

	<u>Page</u>
2.1 This figure shows the behaviour of a gravitational wave that propagates perpendicularly to this paper when it crosses matter. The polarization h_+ is represented in the upper level and in lowest level is h_\times	11
2.2 The first gravitational wave detection: GW150914.	14
2.3 The two LIGO sites: (left) Hanford, near Washington and (right) Livingston in Louisiana.	16
2.4 The 52 gravitational wave events separated by their sources: black holes and/or neutron stars.	17
2.5 The gravitational wave interferometer's configuration. a) presents the location of both LIGO sites (H1 and L1) and its orientations. b) shows the curve of amplitude spectral density of the instrument noise as a function of the frequency for both sites.	18
2.6 The VIRGO Interferometer located at Santo Stefano a Macerata, Cascina, Italy.	19
3.1 Three sine waveforms are to be analyzed. The parameters used in this plot were: $N=1024$ samples, $T = 8s$, $rate = 128Hz$	22
3.2 Time-Domain Analysis: The waveforms are combined in a sum, just as a detector would do.	23
3.3 Frequency-Domain Analysis: The waveforms are represented individually in frequency-domain as the grey lines, and the red line corresponds to the waveforms combined in a sum.	24
3.4 An arbitrary signal represented in time-frequency domain (left) and the details of the Heisenberg box representing a part of the signal (right). . .	25
3.5 Noise sources from LIGO detectors. The low frequency noises affects mostly the Livingston detector.	28
3.6 Data processing steps performed by the LIGO-Virgo detectors.	29
4.1 The information in Equation (4.1) is represented in time-frequency plane where it depends on the spread (ϵ_γ and u_γ) of the atom ϕ_γ	33
4.2 This is an example of a signal wavelet decomposition of 3 levels. The original data is expressed in the first upper bin (level zero) and while the decompositions are being performed, the bins are divided into low and high frequency. All the bins can be identified with its respective decomposition level l and frequency layer i	35

4.3	AVIRIS image and the results for the wavelet analysis.	36
4.4	Ground truth colours.	37
4.5	Comparison of Wavelet filters.	37
5.1	Data analysis steps performed in this project starting from acquiring the data (input) and ending with the time-frequency data (output) plot. The dashed blocks represents optional steps that the user can choose to use it or not. The Inverse Discrete Wavelet Transform is executed for last in order to verify if the code's output returns to the input.	39
5.2	The GW150914 (H1) data expressed in time-domain, where $t = 0$ represents the moment the event occurs. The parameters used here are $rate = 4096\text{Hz}$, $N = 32768$ samples and $T = 8\text{s}$	41
5.3	The Tukey Window as a function of time, when analyzed with the shape parameters $p = 0$ (Rectangular Window), $p = 0.3$ and $p = 1$ (Hann Window).	42
5.4	The GW150914 (H1) windowed data expressed in time-domain when using a Tukey Window of shape parameters (p) 0.1 and 0.3.	43
5.5	Power Spectral Density of the Noise as a function of the frequency.	44
5.6	The GW150914 (H1) windowed and whitened data expressed in time-domain.	44
5.7	The GW150914 (H1) event whitened and whitened-bandpassed (Filter 1) data comparison in frequency-domain.	45
5.8	The GW150914 (H1) event whitened and whitened-bandpassed data expressed in time-domain.	46
5.9	A quantile example using a simple gaussian distribution.	48
5.10	The results for GW150914 (H1) when evaluated using $T = 2\text{s}$ and a sampling rate of 4kHz. There are two time-frequency plots, the left figure representing the signal when no bandpass filter is applied and the right figure displays the event using the [20,1000]Hz filter. In both plots the event's $f(t)$ is presented as the blue curve.	53
5.11	The results for GW150914 (H1) when evaluated using $T = 8\text{s}$ and a sampling rate of 4kHz. There are two time-frequency plots, the left figure representing the signal when no bandpass filter is applied and the right figure displays the event using the [20,1000]Hz filter. In both plots the event's $f(t)$ is presented as the blue curve.	53

5.12	The GW170817 data with the glitch detected in the Livingston detector represented in time-frequency using $T = 4\text{s}$ and a sampling rate of 4kHz. The plot covers a time interval of $[-1.1,-0.95]\text{s}$ before the event, with 0 seconds being the gravitational wave emission. The glitch is centered in -1.02 seconds and the colorbar scale is defined from the maximum coefficient generated during the wavelet decomposition.	55
5.13	The results for the GW190521 event represented in time-frequency when using $T = 2\text{s}$ and a sampling rate of 4kHz. On the left, there is the time-frequency plot when no bandpass filter is applied and on the right, Filter 1 was used ($[20,1000]\text{Hz}$). In both plots the event's $f(t)$ is presented as the blue curve.	56
5.14	Time-frequency representations for the GW190521_074359 event using $T = 2\text{s}$ and a sampling rate of 4kHz. On the left, there is the time-frequency plot when no bandpass filter is applied and on the right, Filter 1 was used ($[20,1000]\text{Hz}$). In both plots the event's $f(t)$ is presented as the blue curve.	57
5.15	Time-frequency representations for the GW170104 (H1) event using $T = 2\text{s}$, a sampling rate of 4kHz, with no bandpass filter and with different colorbar scales. The right figure will be compared with the other events in order to compare their intensities. In both plots the event's $f(t)$ is presented as the blue curve.	58
5.16	Time-frequency representations for the GW170104 (H1) event using $T = 2\text{s}$, a sampling rate of 4kHz and a bandpass filter of $[20,1000]\text{Hz}$ and with different colorbar scales. The right figure will be compared with the other events in order to compare their intensities. In both plots the event's $f(t)$ is presented as the blue curve.	59
5.17	Time-frequency representations for the noise 200s before the GW150914 event, using $T = 8\text{s}$, a sampling rate of 4kHz and a bandpass filter of $[20,1000]\text{Hz}$	60
5.18	Inverse Discrete Wavelet Transform for the GW150914 (H1) coefficients. The difference between both original and IDWT data is represented with the <i>diff</i> curve.	61
5.19	Inverse Discrete Wavelet Transform for the GW150914 (H1) coefficients filtered with quantile and a bandpass filter ($[20,1000]\text{Hz}$). The difference between both original and IDWT filtered data is represented with the <i>diff</i> curve.	61

5.20	Histogram of the GW150914 (H1) event’s coefficients module. The quantile representing 0.90 and 0.95 of the coefficients curve is described as the green and red lines, respectively.	63
5.21	Histogram of the GW150914 (H1) data’s coefficients module. The coloured boxes indicates the correspondent coefficient in the histogram to the the time-frequency representation tile. The bigger time-frequency plot represents all the coefficients below $ 0.13 $ and its correspondent in the histogram. The quantile representing 0.90 and 0.95 of the coefficients curve is described as the green and red lines, respectively.	64
5.22	Histogram of the GW190630_185205 (H1) event’s coefficients module. The quantile representing 0.90 and 0.95 of the coefficients curve is described as the green and red lines, respectively. The highest coefficient is the only one that can be identified and it is presented with the yellow box.	65
5.23	Final mass analysis.	66
5.24	SNR analysis.	67
5.25	Mass-SNR analysis.	68
5.26	Mass Ratio-SNR analysis.	68
A.1	Time-frequency representations for the GW150914 event using 2 seconds and a sampling rate of 16kHz. The left figure shows the time-frequency plot without filtering and the right figure presents the result for the wavelet representation when using a $[20,1000]$ Hz bandpass filter. In both plots the event’s $f(t)$ is presented as the blue curve.	77
A.2	Time-frequency representation of the GW190408_181802 (H1) data using $T = 8$ s, a Tukey Window shape parameter 0.1, no bandpass filter and a sampling rate of 4kHz. The left figure presents the maximum coefficients corresponding to $ 0.8 $, but since a comparison between the events intensity is made, a plot with the colorbar of the other events is shown in the right figure. In both plots the event’s $f(t)$ is presented as the blue curve.	78
A.3	Time-frequency representation of the GW190408_181802 (H1) data using $T = 8$ s, a Tukey Window shape parameter 0.1, a bandpass filter ($[20,1000]$ Hz) and a sampling rate of 4kHz. The left figure presents the maximum coefficients corresponding to $ 0.8 $, but since a comparison between the events intensity is made, a plot with the colorbar of the other events is shown in the right figure. In both plots the event’s $f(t)$ is presented as the blue curve.	79

A.4	The time-frequency representations for the GW190519_153544 (L1) event, when using $T = 8$ s, a Tukey window shape parameter of 0.3 and a sampling rate of 4kHz. On the left figure there is the result for this event evaluated without a bandpass filter and on the right the result for the event when Filter 1 ([20,1000]Hz) is applied in the data processing. In both plots the event's $f(t)$ is presented as the blue curve.	80
A.5	The time-frequency representations for the GW190630_185205 (L1) event, when using $T = 8$ s, no bandpass filter a sampling rate of 4kHz. On the left figure there is the result for this event evaluated without a bandpass filter and on the right the result for the event when Filter 1 ([20,1000]Hz) is applied in the data processing. In both plots the event's $f(t)$ is presented as the blue curve.	81
A.6	The time-frequency representations for the GW190630_185205 (L1) event, when using $T = 8$ s, a bandpass filter ([20,1000]Hz), a sampling rate of 4kHz and different colorbar scales. In both plots the event's $f(t)$ is presented as the blue curve.	82
A.7	The time-frequency representations for the GW190412 (L1) event, when using $T = 4$ s; without a bandpass filter; a Tukey window shape parameter of 0.3, a sampling rate of 4kHz and different colorbar scales. In both plots the event's $f(t)$ is presented as the blue curve.	83
A.8	The time-frequency representations for the GW190412 (L1) event, when using $T = 4$ s, a bandpass filter ([20,1000]Hz), a sampling rate of 4kHz and different colorbar scales. In both plots the event's $f(t)$ is presented as the blue curve.	83
A.9	The time-frequency representations for the GW190701_203306 (L1) event, when using $T = 8$ s, with a bandpass filter ([20,1000]Hz) and a sampling rate of 4kHz. In the plot the event's $f(t)$ is presented as the blue curve.	84
A.10	The time-frequency representations for the GW190701_203306 (L1) event, when using $T = 16$ s, with a bandpass filter ([20,1000]Hz) and a sampling rate of 4kHz with different colorbar scales. In both plots the event's $f(t)$ is presented as the blue curve.	85

LIST OF SYMBOLS

N	–	number of samples: number of data points
T	–	time interval: total data interval
$rate = \frac{N}{T}$	–	sampling rate/sampling frequency: number of samples per second
$\Delta t_{min} = \frac{1}{rate}$	–	time resolution: minimum distance between two consecutive time bins
$\Delta f_{min} = \frac{1}{T}$	–	frequency resolution: minimum distance between two consecutive frequency bins
$f_{NYQ} = \frac{rate}{2}$	–	Nyquist frequency: the maximum frequency the signal can assume

CONTENTS

	<u>Page</u>
1 INTRODUCTION	1
1.1 Objectives	3
2 GRAVITATIONAL WAVES	5
2.1 General Relativity	5
2.2 Gauges transformations	7
2.2.1 TT gauge	8
2.2.2 Polarization modes	10
2.3 Gravitational waves signals	11
2.4 Science Impact of the Gravitational Wave Detectors	15
2.4.1 Laser Interferometer Gravitational-wave Observatory (LIGO)	15
2.4.2 VIRGO Interferometer	18
3 DATA ANALYSIS BACKGROUND	21
3.1 Time-domain	21
3.2 Frequency-domain	22
3.3 Uncertainty principle	25
3.4 Noise spectral density	26
4 WAVELET THEORY	31
4.1 Wavelets	31
4.1.1 Wavelet transforms	33
4.1.2 Wavelet decomposition	34
4.2 Applications	35
5 ADAPTIVE MULTIREOLUTION FOR WAVELET ANALYSIS 39	
5.1 Methodology	39
5.1.1 Data	39
5.1.2 Windowing	41
5.1.3 Whitening	43
5.1.4 Bandpass filter	45
5.2 Discrete wavelet transform	46
5.2.1 Shannon's entropy criterion	47

5.2.2	Quantile threshold	48
5.3	Results: time-frequency representations	49
5.3.1	GW150914	52
5.3.2	GW170817 (glitch)	54
5.3.3	GW190521	55
5.3.4	GW190521_074359	56
5.3.5	GW170104	58
5.3.6	Noise	59
5.4	Inverse discrete wavelet transform	60
5.5	Analysis	62
5.5.1	Histogram	62
5.5.2	Final mass	66
5.5.3	Signal-to-Noise Ratio	66
5.5.4	Mass-SNR	67
6	CONCLUSION	69
	REFERENCES	73
	APPENDIX A	77
A.1	GW150914 - 16kHz	77
A.2	GW190408_181802	78
A.3	GW190519_153544	79
A.4	GW190630_185205	80
A.5	GW190412	82
A.6	GW190701_203306	84

1 INTRODUCTION

The human race is always chasing the unknown, specially when it comes to understanding and discovering the universe. With a lot of hard work of scientists all over the world, today we are capable of appreciating their effort and to keep pursuing the solutions for the problems that remains.

One of these great scientists was Albert Einstein. When he developed The General Relativity Theory in 1915, it unleashed a huge number of results and solutions for problems that the Newtonian theory wasn't capable of answering. Also, predictions were made with that theory, such as the existence of gravitational waves ([ABBOTT et al., 2016](#)).

By that time, since the scientists made a huge development on observing the universe using the electromagnetic radiation (visible light, radio waves, X-rays, etc), it was logical to treat the gravitational wave (GW) as an analog of this radiation. Therefore, two groups were created: the *skeptics*, led by Herman Bondi, represented the ones who really doubted that gravitational waves exists and *non-skeptics*, led by John Wheeler and represented the ones who believe in it and relied in this analogy.

For many years, this discussion seemed to be endless, until the Chapel Hill Conference in 1957 ([THE... , 1957](#)). Therefore the pioneer experimental effort of Joseph Weber, who built the now known as the *Weber Bar Experiment* ([CHO, 2016](#)), started the race for the detection of gravitational waves, encouraging the development of other detection methods, such as the cryogenic resonator and the laser interferometer, this last one being the most efficient way of detection.

Not until 50 years from Weber's experiments, the first gravitational wave was detected by the Laser Interferometer Gravitational-Wave Observatory (LIGO) ([ABBOTT et al., 2009](#)), as it will be seen in Chapter 2.

What is interesting about studying gravitational waves is that it carries energy and information about its source, so that means, it's possible to measure this "ripples" that are propagated in all directions in space-time and produced by an extremely energetic process.

This is a different way of approaching astrophysical objects and also to study the universe and its origin. But, the results of these detections must be carefully checked.

The results of the detections can be found in the LIGO website, the raw data are

expressed as a time series of reconstructed metric perturbation, but it can also be transformed to Fourier-domain (frequency-domain). The transition and relation between time-domain and frequency-domain are explored in Chapter 3.

There is another option, which is to represent time series in time-frequency domain. The Wavelet Transform (MALLAT, 2009) is a method of performing that task and the mathematical form for this technique is quite similar to the Fourier Transform, it represents an expansion in a base that depends on a time function called *mother wavelet*.

Every mother wavelet is localized in time and frequency, but can also be translated, stretched and contracted without changing its area, so that it can fill the entire time-frequency plane, describing the signal in "tiles" with width given by the time-resolution and height described with frequency-resolution. Chapter 4 will discuss the wavelet method and its applications.

When using the wavelet method, the input is given by a list of coefficients, each of which refers to a time bin and a frequency bin. The signal in wavelet domain is represented as tiles with the same area, but different shapes, adjusting itself according to the signal that is analyzed by the code.

In this work a code in Python is developed using an optimal criterion to choose the best decomposition, or shape of the tiles for the decomposition, given using Shannon's Entropy, that is, calculating this quantity for each pair of coefficients and comparing with the entropy of the layer before. The objective of this criterion is minimizing the entropy value, which physically means representing data points in the way that concentrates the intensity in the lowest number of tiles.

The resulting representation is multiresolution in time and frequency (STURANI; TERENZI, 2007), corresponding to different time and frequency resolutions of the Heisenberg boxes associated to the signal. The area of this box is constant, but with the flexibility of changing the resolutions, the representation adjusts according to the data used as input for the code. This tool is applied to 10 gravitational wave events. The code sketch is described in details in Chapter 5, as well as the results of the time-frequency representation of the GW events and the criterion used to determine the goodness of the representation.

Chapter 6 presents the conclusion of the developed method in this dissertation and some final remarks.

1.1 Objectives

This dissertation aims to develop a tool that generates a multiresolution representation that adjusts to any time series used as input for the wavelet analysis. More specifically:

- To develop a code in Python that processes the data and performs a Discrete Wavelet Transform (DWT) using a selection criterion given by Shannon's Entropy for the decompositions. To plot a multiresolution representation of the data in time-frequency with the outputs of the DWT. Apply these steps to 10 GW events and validate the results.
- Identify the GW parameters such as Signal-to-Noise Ratio, length of the signals, etc, for which the developed method is most accurate.

2 GRAVITATIONAL WAVES

2.1 General Relativity

The theory of General Relativity ([EINSTEIN, 1907](#)) enabled important accomplishments in physics, such as the prediction of gravitational waves (GWs) by Einstein in 1918. In order to achieve a proper physical and mathematical definition of GWs, it is important to have some concepts in mind, such as the Einstein Field Equation:

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu}. \quad (2.1)$$

where $R_{\mu\nu}$ is the Ricci curvature tensor, $g_{\mu\nu}$ is the spacetime metric, G is the gravitational constant and $T_{\mu\nu}$ is the energy-momentum tensor (see ([D'INVERNO; HARVEY, 1993](#)) for more details).

When this equation was solved through different methods by great scientists such as Karl Schwarzschild, Roy Kerr, it resulted in solutions that provided the theoretical existence of astrophysical objects, such as black holes, wormholes, rotating black holes, given the symmetry of the space-time (spherical, axial). Nevertheless, to derive a solution that proves the existence of gravitational waves, the approach needed is slightly different.

It is necessary to use a 'linearized theory' to describe the spacetime metric $g_{\mu\nu}$. From this, $g_{\mu\nu}$ describes a small deviation from Minkowski's flat metric $\eta_{\mu\nu} = \text{diag}(-, +, +, +)$ as the following:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad (2.2)$$

with $h_{\mu\nu}$ representing a perturbation of the spacetime metric, given that $|h_{\mu\nu}| \ll 1$, which implicitly conserves the weak field condition and constrains the coordinate system to be approximately cartesian, so $h_{\mu\nu}$ is a tensor under Lorentz transformations but not under general coordinate transformations. It is also considered here that higher order terms of this perturbation are discarded.

It's possible to evolve this linearized theory by using the Einstein's Equation (2.1) with all the fixed conditions. The starting point is to define the components of the affine connection (Christoffel symbols ([HEINBOCKEL, 1996](#))) given by

$$\Gamma^\mu{}_{\nu\rho} = \frac{1}{2}\eta^{\mu\sigma} (\partial_\rho h_{\sigma\nu} + \partial_\nu h_{\sigma\rho} - \partial_\sigma h_{\nu\rho}) = \frac{1}{2} (\partial_\rho h_{\mu\nu} + \partial_\nu h_{\mu\rho} - \partial_\mu h_{\nu\rho}), \quad (2.3)$$

where ∂_μ represents the partial derivative $\partial/\partial x^\mu$.

This leads to the construction of the Riemann tensor with the components of the affine connection, as it is done next:

$$R_{\mu\nu\rho\sigma} = \frac{1}{2} (\partial_\nu \partial_\rho h_{\mu\sigma} + \partial_\mu \partial_\sigma h_{\nu\rho} - \partial_\mu \partial_\rho h_{\nu\sigma} - \partial_\nu \partial_\sigma h_{\mu\rho}), \quad (2.4)$$

and from the equation above, the Ricci tensor can be written as

$$R_{\mu\nu} = R^\rho{}_{\mu\rho\nu} = \frac{1}{2} (\partial_\rho \partial_\nu h^\rho{}_\mu + \partial_\rho \partial_\mu h^\nu{}_\rho - \square h_{\mu\nu} - \partial_\rho \partial_\mu h), \quad (2.5)$$

with the D'Lambertian operator representing the spacetime components of the wave ($\square = \nabla^2 - \frac{1}{c^2} \partial_t^2$) and $h^\mu{}_\mu$ standing for the trace of the metric perturbation.

Furthermore, the Ricci scalar, that is, the Ricci tensor contraction is defined as

$$R = (\partial_\rho \partial^\mu h^\rho{}_\mu - \square h) \quad (2.6)$$

allowing the construction of the Einstein equation expressed in terms of the metric perturbation. This procedure will be key to the understanding of gravitational waves.

A way of simplifying it is to work with the association $\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h$, where \bar{h} is denominated *trace-reversed* perturbation reduces the expression above to a much simpler equation.

This procedure was developed by Einstein in 1916. For further deductions, see (DIRKES, 2018). Notice that in General Relativity, the right side of the equation above is the energy-momentum tensor $T_{\mu\nu}$ and it can be related to the Einstein tensor $G_{\mu\nu}$ with $G_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}$, that way the field equation is defined as

$$\boxed{G_{\mu\nu} = \frac{1}{2} (\partial_\rho \partial_\nu \bar{h}^\rho{}_\mu + \partial^\rho \partial_\mu \bar{h}_{\nu\rho} - \square \bar{h}_{\mu\nu} - \eta_{\mu\nu} \partial_\rho \partial^\sigma \bar{h}^\rho{}_\sigma)}. \quad (2.7)$$

It is clear that this expression is too complex to work with, so a couple of changes has to be made in order to simplify it. In all physical systems, identifying symmetries are an efficient way of decreasing its degrees of freedom, as it is shown next.

2.2 Gauges transformations

Symmetries are essential when it comes to physics, a successful theory respects conservation laws such as Noether's and can be locally analysed by different configurations. These configurations are related to each other through "gauge transformations" and it is a procedure made in order to restrain the physical's degrees of freedom, as it's done in Maxwell's electromagnetism theory (GRIFFITHS, 2010). In General Relativity this concept is no different, gauges transformations represents a coordinate transformation. So, the gravitational wave theory must be now rewritten, Equation (2.7) more specifically, so physical aspects can be determined and a better understanding of this phenomenon can be made.

First, an infinitesimal coordinate transformation $x'^{\mu} = x^{\mu} + \epsilon^{\mu}$ is defined, which ϵ^{μ} represents an infinitesimal displacement, so the metric becomes:

$$g'_{\mu\nu} = g_{\mu\nu} - \partial_{\mu}\epsilon_{\nu} - \partial_{\nu}\epsilon_{\mu}. \quad (2.8)$$

With this gauge freedom, it is wise to choose an approximate inertial frame of reference to solve the D'Alembert's equation, after all, the cartesian coordinates are used for that purpose in special relativity (NUSSENZVEIG, 2014). A convenient gauge choice would be **De Donder gauge** $g^{\alpha\beta}\Gamma^{\gamma}_{\alpha\beta} = 0$ or the linearized version (analogous to **Lorentz gauge**):

$$\partial^{\nu}\bar{h}_{\mu\nu} = 0. \quad (2.9)$$

When the gauge is applied to Einstein tensor (Equation (2.7)), all that's left is the following:

$$\square\bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4}T_{\mu\nu}, \quad (2.10)$$

since $G_{\mu\nu}$ can be expressed in terms of $T_{\mu\nu}$, the equation above represents a wave, in this case a gravitational wave $\bar{h}_{\mu\nu}$ and the solution for it are given by a Green's function. More preciously, a retarded Green's function in three dimensions, analogous to the potential of a moving charge in electrodynamics (GRIFFITHS, 2010):

$$\bar{h}_{\mu\nu}(t, \vec{x}) = \frac{4G}{c^4} \int T_{\mu\nu} \left(t - \frac{|\vec{x} - \vec{x}'|}{c}, \vec{x}' \right) \frac{1}{|\vec{x} - \vec{x}'|} d^3x' \quad (2.11)$$

written in cartesian coordinates. A particular solution for this equation is considering

a system very far from the source, so $T_{\mu\nu} = 0$, it comes to

$$\square \bar{h}_{\mu\nu} = 0. \quad (2.12)$$

Since $\square = \nabla^2 - \frac{1}{c^2} \partial_t^2$, this wave is being propagated at the speed of light, and the plane wave solution have the form:

$$\bar{h}_{\mu\nu} = A_{\mu\nu} e^{ik_\alpha x^\alpha} \quad (2.13)$$

the polarization tensor indicated by $A_{\mu\nu}$ contains information about the amplitude and like the name says, the polarization of the wave. The wave vector is expressed in terms of the frequency w and the speed of light c : $k^\alpha = (\frac{w}{c}, \vec{k})$. It's important to remember that these quantities are all constants and since it's also imaginary, only the *real* component of Equation (2.13) is used when the analysis are made.

The only gauge used until now is Lorentz's, which restrained 10 degrees of freedom of the gravitational wave tensor $\bar{h}_{\mu\nu}$ into 6 degrees, so another attempt of limiting the degrees of freedom even more will be made through the so-called **TT gauge**.

2.2.1 TT gauge

A condition for the polarization tensor is assigned from De Donder gauge (written as Equation (2.9)) when the derivation of (2.13) is done:

$$A_{\mu\nu} k^\mu = 0 \quad (2.14)$$

since the scalar product of the polarization and the wave vector is zero, it implies that these vectors are orthogonal to each other. Also, another statement can be made: the number of independent components of $A_{\mu\nu}$ are reduced to 4.

An analysis of the wave vector can be done taking into consideration the condition above. From the scalar product definition:

$$k_\mu k^\mu = k_0^2 - k_x^2 - k_y^2 - k_z^2 = 0, \quad (2.15)$$

so it means the wave vector k_μ is lightlike or null, confirming that the wave \bar{h} travels at the speed of light and frequency w .

Until now, the wave amplitude had originally 10 components, with Lorentz gauge

this number was restrained to 4 components. So this is where the TT gauge is applied in vacuum it will reduce this number even more, with the following procedure.

Equation 2.9 rewritten with $\bar{h}_{\mu\nu} = h_{\mu\nu}$ when $\mu = 0$ and with a fixed h_{0i} is:

$$\partial^0 h_{00} = 0 \quad (2.16)$$

which physically means that the h_{00} component is the static part of the GW and since the gravitational wave is time dependent, $h_{00} = 0$. All that is left are the spatial components represented by $h_{0\mu} = 0$ and when Lorentz gauge is applied, the condition changes into $\partial h_{ij} = 0$. Since the transverse-traceless gauge is applied, the condition becomes:

$$h_i^i = 0. \quad (2.17)$$

The objective of these conditions are to reduce the degrees of freedom and after these procedures there are 2 degrees of freedom left, leaving two polarization modes for the gravitational wave.

This gauge provides an identity between the metric perturbation ($h_{\mu\nu}$) and the gravitational field ($\bar{h}_{\mu\nu}$):

$$\bar{h}_{\mu\nu} = h_{\mu\nu} \quad (2.18)$$

and for a better understanding, the wave vector under the TT gauge is defined as $h_{\mu\nu}^{TT}$.

When resuming to how this gauge will affect the components number of the amplitude $A_{\mu\nu}$ tensor, symmetries are determined and this tensor will be described as 2 dimensionless amplitudes h_+ and h_\times :

$$\boxed{A_{\mu\nu} = h_+ \epsilon_+^{\mu\nu} + h_\times \epsilon_\times^{\mu\nu}}, \quad (2.19)$$

with the amplitude $A_{\mu\nu}$ written as a linear combination of both components and described with the *unit polarization tensors* $\epsilon^{\mu\nu}$ given by

$$\epsilon_+^{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$\epsilon_\times^{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

this way it is possible to define the gravitational wave amplitudes, that is, the way the wave moves from a source and crosses the universe. For a further discussion on the subject, see (MAGGIORE, 2009).

2.2.2 Polarization modes

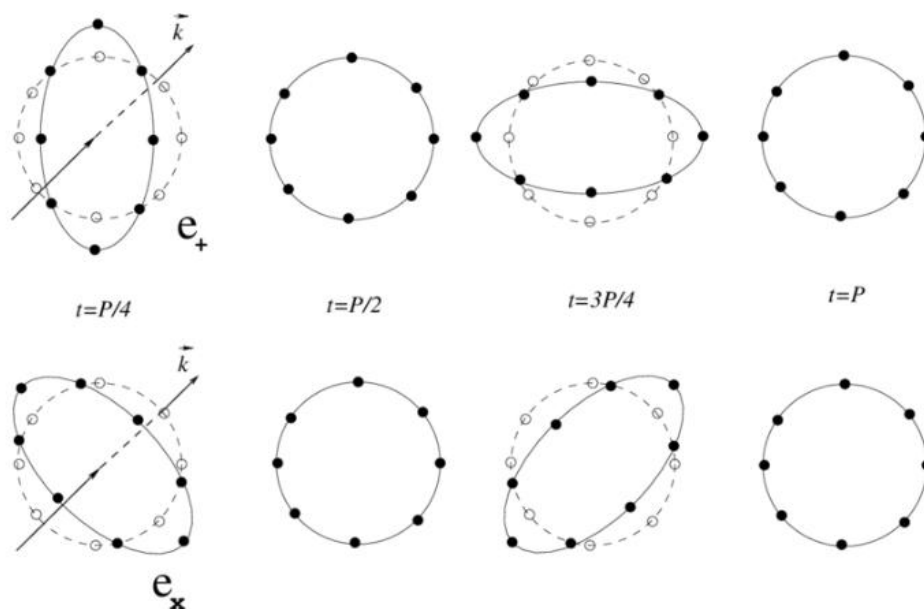
When it comes to gravitational radiation, that is, the emission from an astrophysical source and propagation through a particle (graviton) or by a gravitational wave, its polarization is essential to understanding where it came from and how similar this ripples in space-time are to the electromagnetic radiation.

Resuming to the electromagnetic theory (GRIFFITHS, 2010), when a small source is generating electromagnetic waves, this radiation is mainly dipolar (since there's no magnetic monopole and higher multipolar orders are discarded). As for the gravitational radiation, when the multipolar decomposition is made, the lowest order component is the quadrupole moment, which represents the vibrational modes a gravitational wave possess (for further details, see (MAGGIORE, 2009)).

The next step would be to define these polarization modes for the gravitational waves (CAI et al., 2017). This vibrational modes are represented by the amplitudes h_+ and h_\times that are orthogonal to each other, as it is shown on Figure 2.1.

A question to be asked at this point is: which astrophysical objects and at which physical condition it will emit gravitational waves? The next section will give an approach to answer this matter.

Figure 2.1 - This figure shows the behaviour of a gravitational wave that propagates perpendicularly to this paper when it crosses matter. The polarization h_+ is represented in the upper level and in lowest level is h_\times .



SOURCE: Bishop e Rezzolla (2016)

2.3 Gravitational waves signals

This section is dedicated to the understanding of gravitational wave types generated by specific astrophysical sources.

A GW source needs to generate a significant perturbation in space-time, so it is necessary for this source to be composed of massive objects (one or more), moving at a high velocity and when the parts of this source are accelerated it becomes spherically asymmetric (this last condition is given by the third derivative of its quadrupole momentum (BISHOP; REZZOLLA, 2016)).

There are four categories of gravitational waves signals and the sources that generates them are listed in Table 2.1 and described next in details:

Table 2.1 - Gravitational wave signals.

GW Signals	Examples of sources
Continuous	A rotating neutron star, binary systems far from coalescence
Compact Binary Inspiral	Binary Neutron Star (BNS) Binary Black Hole (BBH) Neutron Star-Black Hole Binary (NSBH)
Stochastic	Big Bang, cosmic strings, superposition of many burst sources
Burst	Impulsive or unmodeled sources

- **Continuous:** This type of gravitational wave signal is well modeled, periodic, and it has a long duration. The sources that generates this type of wave are characterised by a particular final stage of a massive star: a neutron star (NS). As it was said before, only an varying spherically asymmetric object is able to emit a gravitational wave, after all, if that symmetry exists, the space-time curvature will be constant and there wouldn't be any oscillations to be propagated.

In the case of a compact object such as a NS, imperfections in the spherical shape while the star spins will contribute for the variation of the quadrupole momentum. This way, the oscillations caused in space-time will be generated continuously, with constant frequency (equal to twice the NS rotation frequency) and amplitude.

There are some other Continuous GW sources, such as neutron star binaries (BNS), but it's important to emphasize that this objects has to be far from coalescence (merging into one single object).

- **Compact Binary Inspiral:** This is the kind of GW that offers the highest emission incidence detected by the LVK Collaboration. It is a well modeled and short duration signal (in second generation detectors' band). It can be simulated with templates (FIELD et al., 2011) that helps identifying possible GW candidates.

Its sources consists of compact objects such as black hole (BH), neutron star (NS) but in a binary system. This system is composed by two of these astrophysical objects and its coalescence is defined by three stages : *inspiral* (the objects are orbiting one another and moving closer together), *merger* (while the objects gets closer, the distance between them drops, the relative

velocity between them rises, releasing orbital energy in a gravitational wave form) and finally the *ring-down* (which is the final stage of this event, where the merged object keeps oscillating until stops).

The coalescences can occur between stellar blackholes, defined as black holes of masses from 5 to 100 M_{\odot} (M_{\odot} stands for 1 solar mass) that can generate an intermediate blackhole, black holes with masses from 10^2 to 10^{5-6} M_{\odot} as GW190521 that generated an intermediate BH as the final remnant of the coalescence.

In the case of a BNS, the final product will depend on the initial masses of the binary components and the energy released in the whole process.

The first GW detection happened in 2015 and it consisted of a compact binary coalescence composed by two black holes with initial masses of 29 M_{\odot} and 36 M_{\odot} . The coalescence stages and more details about this event are shown on Figure 2.2 and (ABBOTT et al., 2016). Gravitational wave detections such as this, are represented by the following initials: *GW150914* where *GW* indicates "Gravitational Wave", followed by the year, month and day, respectively.

It's important to know that each binary pair produces an unique GW pattern, shorter signals tend to come from more massive objects (for example: BBH) and longer signals explain that the maximum frequency emitted in a coalescence is given by the inverse of the binary total mass.

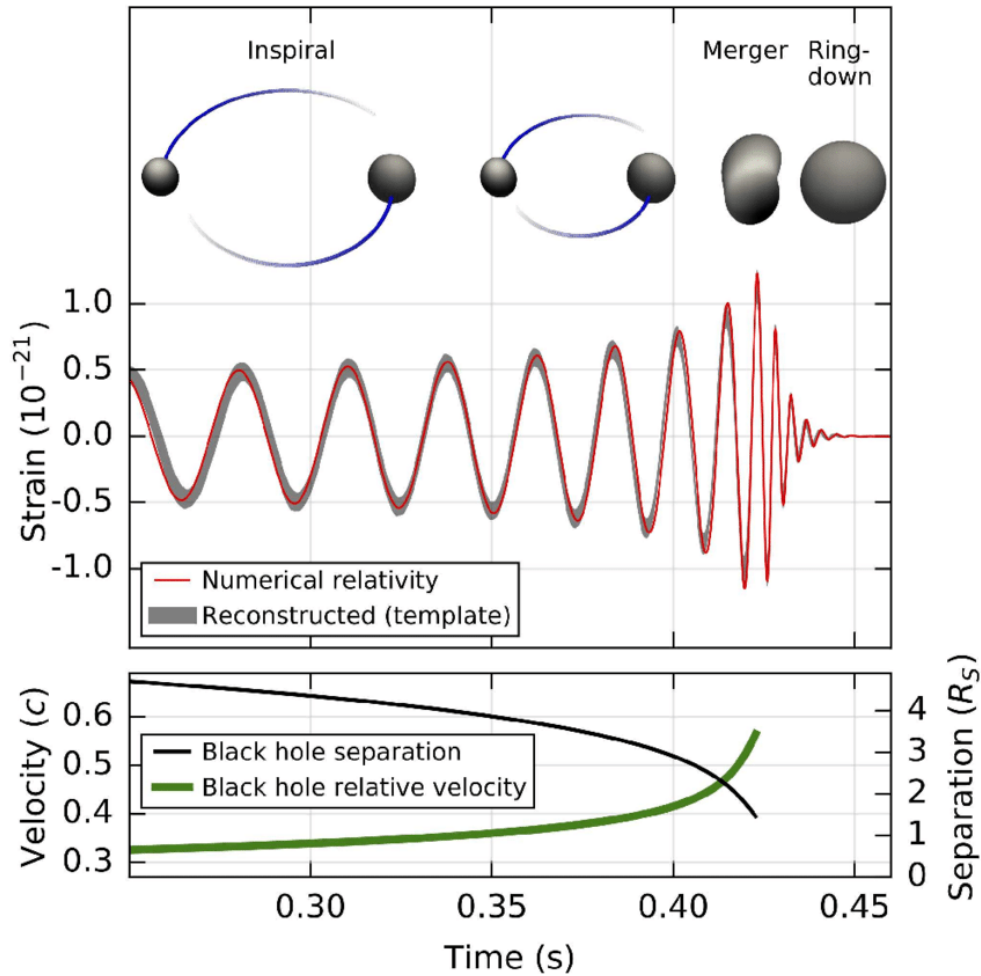
- **Stochastic:** This gravitational wave signal is not modeled and it doesn't have a specific duration, they are rather present all the time.

The sources that produces this type of GW are related to the large number of random, independent events that happens in all directions of the universe. Stochastic gravitational waves can be the key to the understanding of the early universe and its evolution. After all, gravitational waves were emitted at approximately 10^{-36} s to 10^{-32} s after the Big Bang and if detected, it will carry information about the very beginning of the universe!

- **Burst:** This type of GW signal are transients signals in excess of the noise, also not modeled, difficult to simulate. It has a short duration coming from impulsive sources. So, a pattern of signals that hasn't been modeled yet must be recognized, and the sources can be systems not yet seen or predicted. That's why the analyses cannot be restrict anyhow while looking

for the bursts. But its detection is essential to reveal valuable information about the universe.

Figure 2.2 - The first gravitational wave detection: GW150914.



SOURCE: ABBOTT et al. (2016).

All these gravitational wave signals share the same problem: they all need to be extracted from the detector's data. Some of the data analysis techniques will be shown in Chapter 3 for a better understanding of how the noise can interfere with the signal and some filtering methods.

The science behind the gravitational wave detectors such as LIGO and VIRGO is presented in the next section.

2.4 Science Impact of the Gravitational Wave Detectors

The Ligo-Virgo-Kagra (LVK) Collaboration is not just making progress detecting gravitational waves. Their contribution extends to the study of the sources that produces this phenomenon. It unites observations with the electromagnetic radiation such as X-rays, visible light, gamma rays, etc, so different perspectives of the same object can be performed. This results in interactions never seen before, as the BNS event GW170817 (that will be explored later in the results chapter) with their electromagnetic response.

There is also a huge technology development that were improved and created, a few examples are listed next:

- Computation and data analysis;
- High-performance optics;
- Ultrahigh vacuum components and techniques;
- Cryogenic technology;
- Interferometric displacement sensor,

elements that assisted in the creation of the most sensitive instrument ever built.

These technologies can help improve instruments used in another field of science. In the next subsection, a better understanding of the instrumental aspects of the detectors are briefly described.

2.4.1 Laser Interferometer Gravitational-wave Observatory (LIGO)

The next investigation point is to understand how the detection works. Many attempts such as detectors using resonant masses were constructed, such as the brazilian spherical resonant mass detector Mario Schenberg ([OLIVEIRA; AGUIAR, 2016](#)), but eventually the scientists evolved with the interferometer idea of using it as a "antenna" for gravitational wave detection.

The Laser Interferometer Gravitational-wave Observatory was constructed in the late 1990s and started operating in 2003 (to see more details about it, see ([ABBOTT et al., 2009](#))). LIGO operates in two sites in the United States: Livingston and Hanford as it is seen in [Figure 2.3](#).

Figure 2.3 - The two LIGO sites: (left) Hanford, near Washington and (right) Livingston in Louisiana.



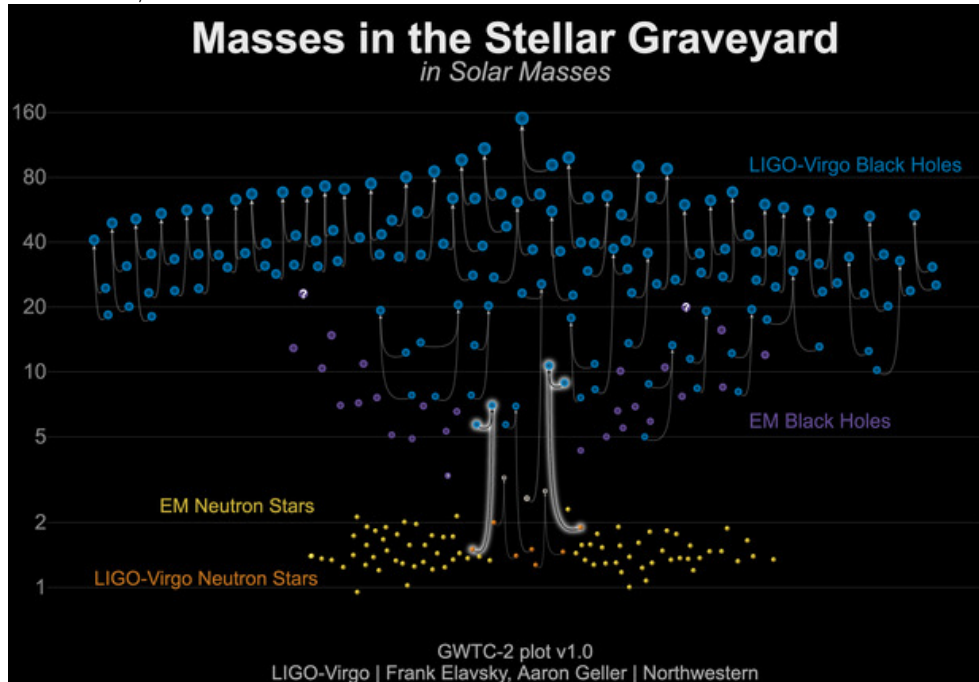
SOURCE: LIGO Caltech (2021).

The search for gravitational waves was made through observation periods, also called "runs":

- **O1**: The first run started on September 12th, 2015 and ended on January 19th, 2016. The results were 3 gravitational wave detections, all of them being generated by black hole mergers (BBH), including the first event GW150914 (ABBOTT et al., 2016);
- **O2**: The second run started on November 30th, 2016 and ended on August 25th, 2017. A total of 8 detections were made, 7 generated by BBH and the first neutron star merger (GW170817 (ABBOTT et al., 2017b)). The Virgo detector (BRADASCHIA et al., 1990) participated in the last month of the O2. Both O1 and O2's catalog of events are presented at (ABBOTT et al., 2019),
- **O3**: The third run is divided into **O3a** that started on April 1st and ended on September 30th, 2019 and **O3b** from November 1st, 2019 and was interrupted on March, 2020 due to COVID-19 pandemic (ROTHAN; BYRAREDDY, 2020). So far, 39 events were detected in O3a as it is seen at (ABBOTT et al., 2021). During the entire run, the VIRGO detector also participated in the GW search.

There were 52 gravitational wave events detected in the three observation runs by the LIGO-Virgo Collaboration so far, and they mark the collisions of two black holes, a black hole and a neutron star, a black hole and a neutron star or two neutron stars, as it is displayed in Figure 2.4.

Figure 2.4 - The 52 gravitational wave events separated by their sources: black holes and and/or neutron stars.



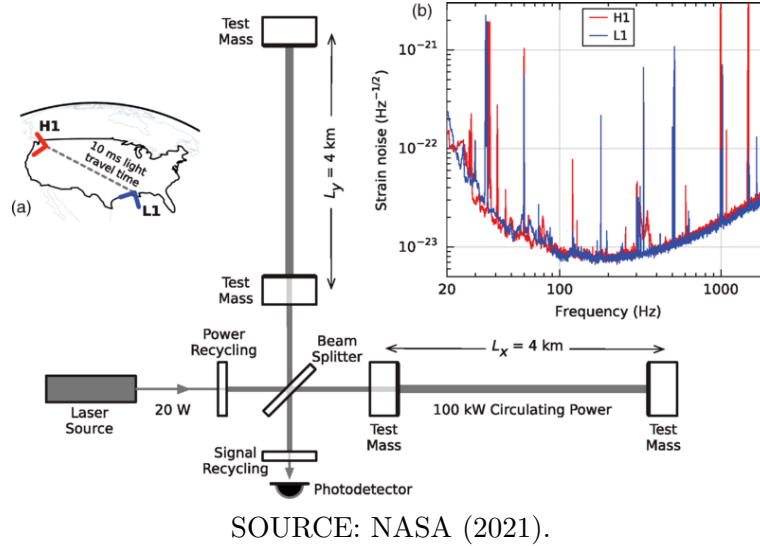
SOURCE: NASA (2021).

The interferometer's layout is presented in Figure 2.5. It is based on the Michelson interferometer, from the Michelson-Morley experiment that was performed in 1887 to detect aether (LANGANGEN et al., 2012). Of course LIGO's arms are a lot bigger (4 kilometers) than Michelson's of 1.3 meters, but they don't differ only in scale, but also with the arms acting as Fabry-Perot cavities (MAGGIORE, 2009).

The GW induces space ripples change the optical path of laser light by that amount, affecting the intensity of light coming out of the interferometer. Therefore all the data generated by this interferometer is evaluated by performing a Windowed Fourier Transform (a subject that will be addressed later) for little chunks of data, generating a spectrogram.

In this configuration the mirrors serve as gravitational test masses, and an advantage in using a resonant Fabry-Perot optical cavity in the arms is that the signal is amplified by a factor of 100 for a 100Hz gravitational wave. By means of that, the LIGO interferometer is considered the most sensible measurement device ever created.

Figure 2.5 - The gravitational wave interferometer's configuration. a) presents the location of both LIGO sites (H1 and L1) and its orientations. b) shows the curve of amplitude spectral density of the instrument noise as a function of the frequency for both sites.



2.4.2 VIRGO Interferometer

Located in Cascina, near Pisa in Italy, the VIRGO Interferometer (see Figure 2.6) was built and started operating in the 2000s (BRADASCHIA et al., 1990). Its configuration is also based on the Michelson interferometer, similar to LIGO's, but with arms 3 meters long.

When the detector was first built, it was not sensitive enough to detect gravitational waves. Therefore, in 2011 it went through improvements and became known as the Advanced Virgo detector with a sensitivity increased by a factor of 10.

VIRGO joined LIGO's run on 2017, participating in the detections ever since the last month of O2. The first detection of this interferometer was GW170814 and it is essential for the sky localisation of gravitational wave signals.

Other gravitational wave observatories are active nowadays, such as GEO600 in Germany (LÜCK, 1997), etc. Plans for a space detectors such as LISA (JOFFRE et al., 2020), TianQin (HU et al., 2018) are being developed, as well as improved third generation earth-based detector as Einstein Telescope (PUNTURO et al., 2010) and others.

Figure 2.6 - The VIRGO Interferometer located at Santo Stefano a Macerata, Cascina, Italy.



SOURCE: Wikipedia (2021).

It is important to remember that when a gravitational wave candidate is identified, this candidate must be detected in at least another gravitational wave detector site. After all, a noise or a glitch can provoke the interference, but the concurring presence of compatible signals in widely separately detectors lower significantly the probability that they are noise artifacts.

The data analysis background used for the detections are presented in the next chapter.

3 DATA ANALYSIS BACKGROUND

The first step of analyzing a signal is to represent the points as a function of an independent variable such as amplitude, time and frequency. In the next sections, two methods are explored: the time-domain and the frequency-domain analysis.

3.1 Time-domain

When data is expressed in time-domain, the time series is a list of amplitude values as a function of time. As the points of a function or a signal changes with time, they can be distributed in a continuous or discrete way. These definitions are very important when defining the parameters of a time-series which will be presented. For discrete data, which will be the ones analyzed in this dissertation, the parameters are shown next:

- N - **number of samples**: number of data points;
- T - **time interval**: total data interval;
- $rate = \frac{N}{T}$ - **sampling rate/sampling frequency**: number of samples per unit of time;
- $f_{NYQ} = \frac{rate}{2}$ - **Nyquist frequency**: the maximum frequency the signal can assume;
- $\Delta t_{min} = \frac{1}{rate}$ - **time resolution**: minimum distance between two consecutive time bins,
- $\Delta f_{min} = \frac{1}{T}$ - **frequency resolution**: minimum distance between two consecutive frequency bins.

Treating data in time-domain can return some very interesting results, such as measuring real-time waves in an oscilloscope. More specifically in astrophysics, a detected signal in every frequency band contains noise (MAGGIORE, 2009). Sometimes one can reach a limit where the signal is completely immersed in noise, therefore a time-domain representation is not useful.

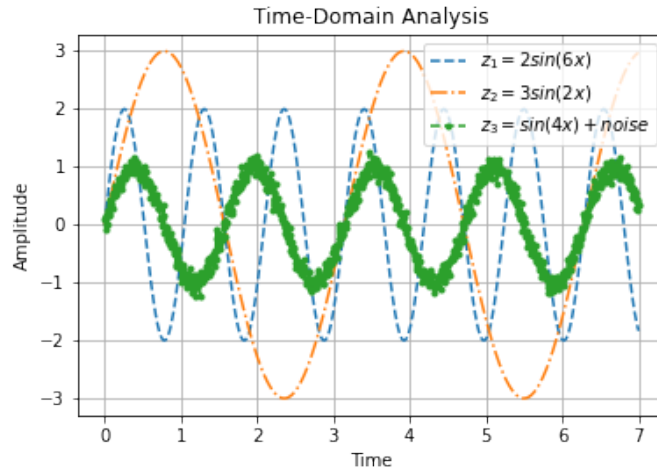
Data can be represented as a function of frequency, that is, in **frequency-domain**, as it is shown next.

3.2 Frequency-domain

Using a frequency-domain analysis, interesting features such as the frequency peaks of the signal, the Power Spectral Density (ABBOTT et al., 2020) can be determined and used in the analysis.

The frequency-domain representation can be understood as another point of view of analyzing the signal, as it will be shown next. In Figure 3.1 there are three sine waveforms, containing different amplitudes and noise in one of them. Let's assume these waves ought to be detected and analyzed.

Figure 3.1 - Three sine waveforms are to be analyzed. The parameters used in this plot were: $N=1024$ samples, $T = 8s$, $rate = 128Hz$.

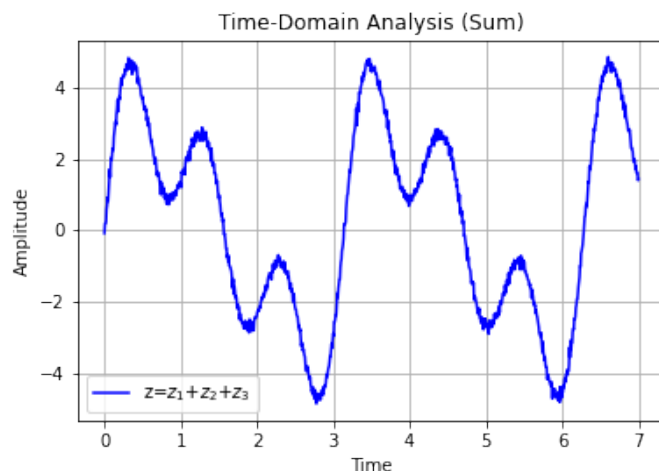


The time and frequency domains are related as it is shown in Figure 3.4, with time being in the x-axis, frequency in y-axis and amplitude in z-axis. Therefore when a time-domain analysis is done, the signal is represented as a sum in time-amplitude plane, as it is shown in Figure 3.2.

Since the time-amplitude plane results are not conclusive, it might be useful to represent the signal in a different way. For that reason, the signal can be represented in Frequency-Domain, or the also called Fourier Domain.

There is a way to convert any data from time-domain to frequency-domain and it is through the Fourier Transform. Consider a function $f(t)$ described in time-domain (it can be a time series describing a GW signal).

Figure 3.2 - Time-Domain Analysis: The waveforms are combined in a sum, just as a detector would do.



This function can be represented in frequency-domain using the following transform, as it is shown next:

$$\tilde{f}(\omega) = \int_{-\infty}^{+\infty} f(t)e^{i\omega t} dt \quad (3.1)$$

where $\tilde{f}(\omega)$ shows how many oscillations per unit of frequency ω are present in the signal $f(t)$ (BUTKOV; SPOSITO, 1969). As long as this new function in frequency can be integrated, an Inverse Fourier Transform can be determined:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega)e^{-i\omega t} d\omega \quad (3.2)$$

the equation above represents the decomposition of the signal $f(t)$ in terms of sinusoidal waves sum ($e^{-i\omega t}$) and if the signal is continuous, its amplitude corresponds to $\tilde{f}(\omega)$.

If the data is discrete, a Discrete form of the Fourier Transform will be used in the analysis. Therefore the Discrete Fourier Transform is defined as:

$$\tilde{f}[\omega] = \sum_{j=0}^{N-1} f[t_j]e^{\frac{i2\pi k}{\Delta t}} \quad (3.3)$$

this identity determines that a sequence of events $f[t_j]$ can be expressed as a sequence of frequency samples.

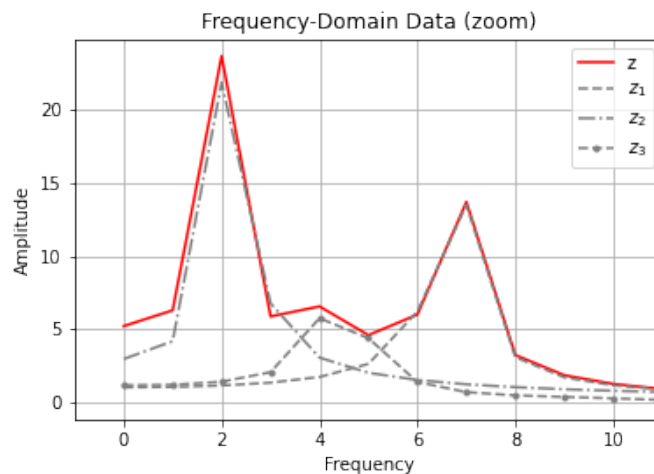
Similar to the Continuous Fourier Transform, an inverse form must exist, but for discrete time, so the inverse form of Equation (3.3) is defined as follows:

$$f[t_j] = \frac{1}{N} \sum_{j=0}^{N-1} \tilde{f}[\omega] e^{-\frac{i2\pi k}{\Delta t}}. \quad (3.4)$$

Assuming that $f[t]$ is the function represented in Figure 3.2 (right), when the Fourier Transform is made, the signal is represented in a frequency-amplitude plane, as it is shown in Figure 3.3. The analyzed function is the sum of all the waveforms, in this domain, it is given as a function of the frequencies. The highest frequency the signal can assume is called the Nyquist frequency (f_{NYQ}) which is given by half of the sampling rate, according to the Nyquist–Shannon sampling theorem (YADAV, 2009).

When comparing the time-domain analysis with the frequency-domain analysis a few conclusions can be identified, such as the waveforms in time-domain cannot be identified individually, they present a pattern, which can be associated with a sinusoidal signal, but it is still a combination of the signals. As for the frequency-domain, it is very clear seeing the frequency peaks that there are three waveforms that has different amplitudes.

Figure 3.3 - Frequency-Domain Analysis: The waveforms are represented individually in frequency-domain as the grey lines, and the red line corresponds to the waveforms combined in a sum.



In conclusion, for this case the Fourier-domain represents the data in a way that the waveforms stands out from the other data points, and even the noise is well defined in amplitudes close to zero.

3.3 Uncertainty principle

As it was presented in these past two sections, time and frequency domains are deeply related, they form not just a 3D space, but also two "points of view" of representing data, although, this connection brings limitations.

Figure 3.4 - An arbitrary signal represented in time-frequency domain (left) and the details of the Heisenberg box representing a part of the signal (right).

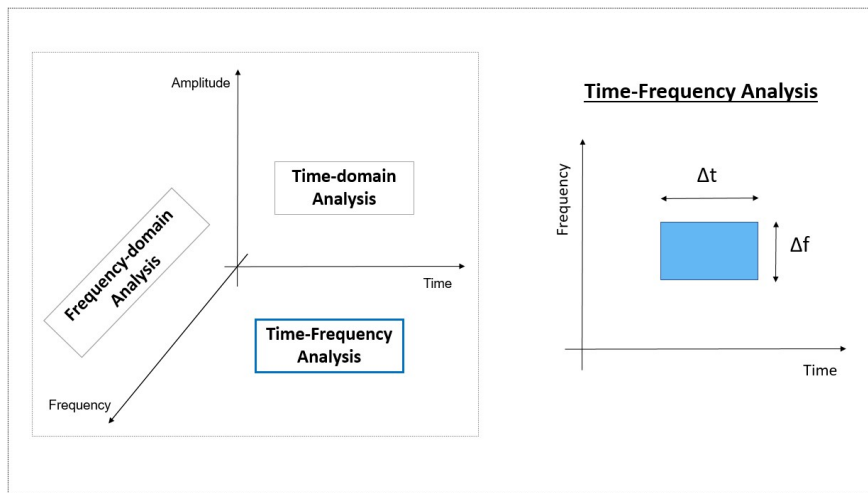


Figure 3.4 shows geometrically how time and frequency are related. Until now, all the analysis with simple functions such as the three waveforms in Figure 3.2 were performed in time-domain or frequency-domain. That very same function can be described in a time-frequency plane, but the difference is that the data will be represented in coloured "tiles". Every tile has a well defined area, with sides expressed by the resolution in frequency Δf and time Δt .

In 1946, Dennis Gabor was the first to treat signals as a quantum system and as a result of this analogy (MACLENNAN, 1994), it became clear that the name of this relation between time and frequency is the Heisenberg-Gabor Uncertainty Principle:

$$\Delta t \Delta f \geq \frac{1}{2} \tag{3.5}$$

where Δt is the time resolution and Δf is the frequency resolution. When performing a wavelet analysis for a signal, the time-frequency representation is given through the Heisenberg Boxes, or also called "tiles" whose area is given by $1/2$. Furthermore, this work offers a multiresolution representation (due to a stopping criterion for the wavelet decompositions), that means, the signal is represented by multiple time and frequency resolutions forming tiles of area $1/2$.

With that flexibility, this allows the tiles to fit in according to the signal used as input: if the tile is very long in frequency, it will be narrow in time, so the area can be respected and so on. To quantify that, the smallest distance between two consecutive time bins is given by $\Delta t_{min} = \frac{1}{rate}$, in the case the associated frequency resolution is the largest: $1/(2\Delta t)$ and when the smallest distance between two consecutive frequency bins $\Delta f_{min} = \frac{1}{T}$ represents the signal, the longest distance in time is presented.

In the Fourier-domain the data has a very accurate frequency resolution, but a very poor one on time, as for the time-domain the frequency resolution is poor, in this case $\Delta f = 1/(2T)$, $\Delta t = T$. Between time-domain and frequency-domain there are many intermediate representations, but this dissertation uses the one that generates a time-frequency representation which concentrates the signal at maximum, using the lowest number of tiles.

The developed code in this work provides the multiresolution representation for the signals, which adapts itself according to the Shannon's entropy values, respecting the area of the tiles and using this flexibility of time and frequency resolutions to give the most refined representation of the input signal in the wavelet domain.

Unfortunately, it is still difficult to identify the event immerse in a huge amount of noise. But there are a couple of filtering tools used in data analysis, such windowing, whitening and applying a band-pass filter will be very useful before any further analysis.

But before that, it is important to understand how a detector such as LIGO captures data that hopefully contains a gravitational wave signal (with noise).

3.4 Noise spectral density

Working with a detector can be quite challenging when it comes to gravitational waves, so setting up tools to detect and extract information from the acquired data is essential (MAGGIORE, 2009). After all, noise is resulted from unpredictable ac-

tions inside and outside the detector and it is reflected in the output measurements. Thereby the input and output will be carefully evaluated with the following procedure.

It will be easier to consider the GW detector as a linear system and so as the relation between signal and noise. Therefore, the first step is to define a mathematical form for the detector's input, as it is done next:

$$h(t) = D^{ij}h_{ij}(t), \quad (3.6)$$

where $h(t)$ represents the input, D^{ij} is denominated as the detector tensor (it is a constant and related to its geometry) and the gravitational wave polarization is denoted by the tensor h_{ij} .

The data generated from the detector is a time series, as it was spoken before, it is a collection of observations listed in time order. This time series is composed by a gravitational wave (hopefully) and noise, as the following:

$$s(t) = h(t) + n(t) \quad (3.7)$$

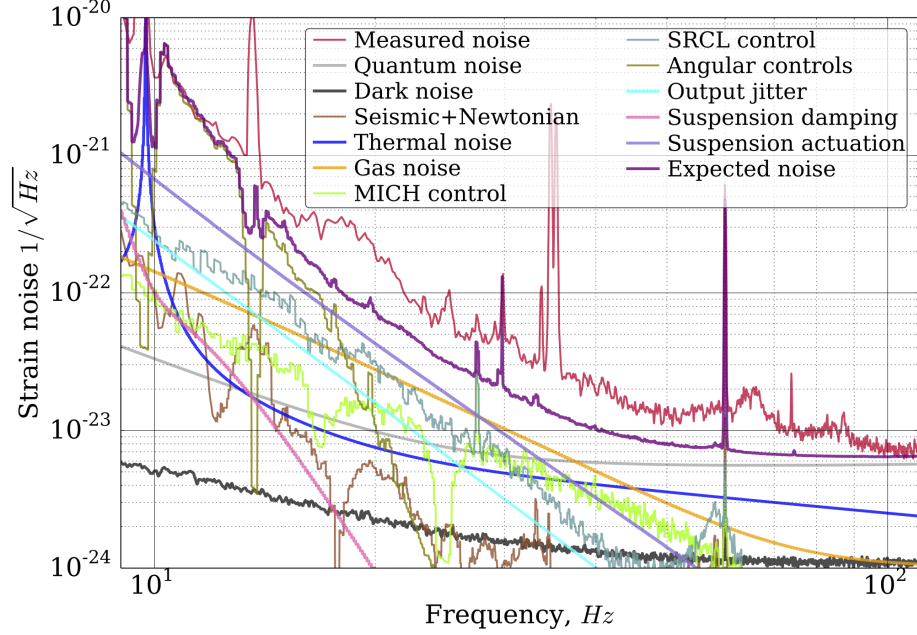
where the GW signal is represented by $h(t)$ and $n(t)$ the noise. The next step is to understand the noise behaviour in the data, therefore it's wise to assume that $n(t)$ is the detector's noise and the only one in the data for now.

Since a spectrum is defined as the statistical average of the signal when analysed in terms of frequencies components, the Power Spectral Density describes the power distribution of a signal along its frequencies, a very useful way to identify any data patterns and further more, the square root of this quantity is called Amplitude Spectral Density, used in the calculation of the amplitude's noise as it is shown next:

Note that in real detections $n(t)$ is a summation of many type of noises, such as quantum, Newtonian, seismic, and even unknown sources, as it is shown in Figure 3.5, where in the y-axis the Amplitude Spectral Density defined as a function of the frequency in the x-axis.

The power spectral density of the noise will be used in the analysis, more specifically in the whitening process, and also to verify if the bandpass filter is being performed correctly.

Figure 3.5 - Noise sources from LIGO detectors. The low frequency noises affects mostly the Livingston detector.



SOURCE: Abernathy et al. (2018).

Due to the physical range of frequencies is defined as $f > 0$, the Noise Spectral Density represented by a discrete form of S_n is being estimated in one *single-side*, that's the reason for the factor $\frac{1}{2}$ to be multiplying S_n as follows:

$$\frac{1}{2}S_n(f) = \langle |\tilde{n}(f)|^2 \rangle \Delta f. \quad (3.8)$$

and in a continuous form:

$$\frac{1}{2}S_n(f)\delta(f - f') = \langle \tilde{n}^*(f)\tilde{n}(f') \rangle. \quad (3.9)$$

where $\tilde{n}(f)$ is the noise data represented in Fourier-domain. Both forms of the above physical quantity is measured in the same unit, as the inverse of frequency $[Hz]^{-1}$.

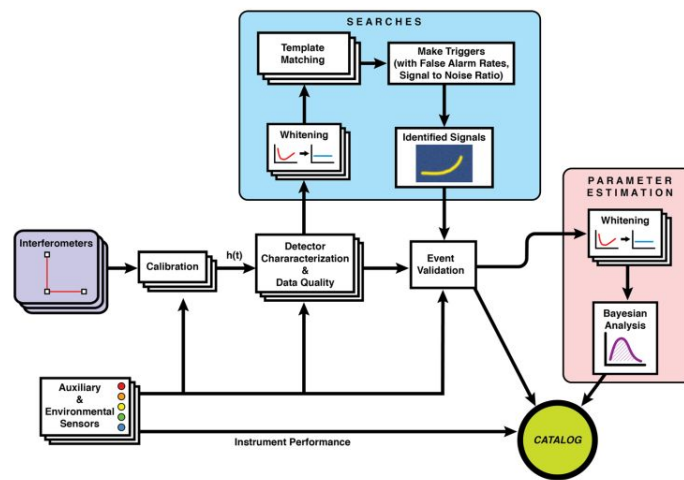
This leads to some physical consequences, such as:

- *White noise*: A random noise that presents the same intensity at different frequencies. It happens when the amplitude is independent of the frequency f , therefore results in a constant noise spectral density function S_n .

- *Colored noise*: A noise in which the amplitude depends on the frequency f , so S_n is not constant anymore, it can vary according to the power-law $S_n(f) = 1/f^\beta$, where $\beta = 0$ (white noise), $\beta = 1$ (pink noise) and so on.

The data analysis steps that LIGO-Virgo follows from the raw data to the detection and validation is presented in Figure 3.6, it is clear that many candidates are identified as glitch or noise as going through all the steps of the analysis.

Figure 3.6 - Data processing steps performed by the LIGO-Virgo detectors.



SOURCE: ABBOTT et al. (2020).

There are other ways of representing a signal, not just a time-domain or frequency-domain, as it was seen on Figure 3.4, with the "tiles". This method will be described in details in the next chapter.

4 WAVELET THEORY

4.1 Wavelets

As it was seen in the previous chapter, when trying to describe a signal using a Fourier approach, it is only possible to extract information of the function in the frequency-domain with a very good resolution in frequency, leaving the time resolution very poor. That means the signal would have the frequencies well determined and the time the frequencies were registered uncertain. Luckily, another attempt to analyze a signal is presented using the "tiles" of the time-frequency representation shown in Section 3.3.

The method that fulfils this "tile representation" is called Wavelets. The idea is to use a function (called *mother wavelet*) (MALLAT, 2009), that is a family of well known waveforms in time and frequency, and correlate it with every data point taken directly from the time series.

In order to perform this task, is it important to fix a time interval of the signal, this is made through a Windowing Fourier Transform. And the family of well defined waveforms was originated in Gabor's studies of energy propagation and are called time-frequency atoms.

A family of time-frequency atoms is represented by ϕ_γ where $\phi_\gamma \in L^2(\mathbb{R})$. A linear transform T of a signal $f(t) \in L^2(\mathbb{R})$ is given by

$$Tf(\gamma) = \int_{-\infty}^{+\infty} f(t)\phi_\gamma^*(t)dt, \quad (4.1)$$

and its frequency form is

$$Tf(\gamma) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega)\hat{\phi}_\gamma^*(\omega)d\omega, \quad (4.2)$$

if $\phi_\gamma(t) \approx 0$ the equations above will only depend on the translations of $f(t)$ in the time-frequency plane.

The objective of this analysis is to build a model for the "wavelet domain" to follow. Therefore, a Windowed Fourier (KEMAO, 2007) can be related to the family of atoms ϕ_γ as the following:

$$\phi_\gamma(t) = g_{\epsilon,u} = e^{i\epsilon t}g(t - u) \quad (4.3)$$

where g represents the time window translated by a quantity u and it is modulated

by a frequency ϵ .

With the fixed relation, the next step is to define a *wavelet atom*, which is a dilation of this translated window by a factor s and relating this quantities with a function ψ called *mother wavelet*, as it is done next:

$$\phi_\gamma(t) = \psi_{s,u}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right). \quad (4.4)$$

The energy expressed by wavelets and Windowed Fourier is well located in time and frequency, which makes it easier for getting a good combination of frequency and time localization for the data. It is different than the Fourier Transform in which the energy is only well located in a limited band of frequencies.

By measuring energy in these atoms, there is a way to relate the energy, the time and frequency in a plot using "tiles", technically the *Heisenberg Boxes* (MALLAT, 2009). It is necessary to define the coordinates that will be used to built this tile. In order to do so, the Uncertainty Principle for signals, also called as Garbor limit is essential in the definition of the resolution in time and frequency.

This is how the Heisenberg box ("tile") is built:

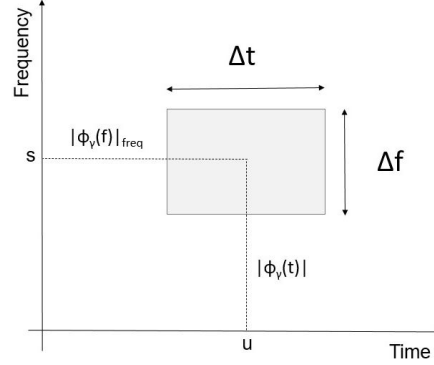
- The box is located in time (x-axis) and frequency (y-axis),
- It is centered in (u, s) ;
- The length along the time axis is $\Delta t(\gamma)$;
- The length along the frequency axis is $\Delta f(\gamma)$,
- The area of the tile is given by:

$$\Delta t \Delta f = \frac{1}{2}.$$

Attending all these restrictions, the Heisenberg boxes are shown in Figure 4.1.

Now that the wavelet domain is established, the next step is to perform the transition between the data points in time-domain to the time-frequency plane.

Figure 4.1 - The information in Equation (4.1) is represented in time-frequency plane where it depends on the spread (ϵ_γ and u_γ) of the atom ϕ_γ .



4.1.1 Wavelet transforms

In order to make signal analysis using wavelets, it is necessary to use all the the wavelet atoms presented before. When the signal goes through a wavelet transform, it gets decomposed in packets over a dilated and translated wavelet (MALLAT, 2009).

First, a wavelet is a function $\psi \in L^2(\mathbb{R})$ with average zero when it is normalized and centered in the neighbourhood of $t = 0$:

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0. \quad (4.5)$$

The next step is to write an equation that represents the wavelet transform, a wavelet operator W will perform this task, as follows:

$$Wf(u, s) = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-u}{s} \right) dt \quad (4.6)$$

where $f(t)$ (time series) is being written directly in wavelet domain (time-frequency representation). That expression can be rewritten using the Convolution Theorem (BUTKOV; SPOSITO, 1969), taking the Linear Filtering into account to achieve the following result:

$$\bar{\psi}_s(t) = \frac{1}{\sqrt{s}} \psi^* \left(\frac{-t}{s} \right), \quad (4.7)$$

and its Fourier transform is given by

$$\hat{\bar{\psi}}(\omega) = \sqrt{s} \psi^*(s\omega). \quad (4.8)$$

All these different forms of expressing Equation (4.6) has the same meaning: it is representing the variation of the signal $f(t)$ in the neighbourhood of u and s , so a dilated and translated wavelet.

What is the difference between the Discrete Fourier Transform (DFT) and the Wavelet Transform (WT)? Well, the DFT uses a time series and perform a transform to the frequency-domain, as for the WT, every data point from the time series are transformed directly into a time and frequency plane. In the DWT case they are written as rectangles with well defined length and height, Δt and Δf respectively.

The next step is to investigate a bit more of how the time series is transformed to the wavelet domain, as it will be seen in the following subsection.

4.1.2 Wavelet decomposition

In signal processing, the resolution is a key point to the analysis. When using wavelets, the resolutions are given using time and frequency. When the signal is represented using wavelets, the wavelet transform's output will have to be carefully filtered so the essential information of the signal can be obtained (MATVEEVICH, 2015).

The wavelet transform's output gives the signal's coefficients (nodes) written in a subspace of level l and layer i defined by \mathbf{W}_i^l . It can be represented as a binary tree as Figure 4.2.

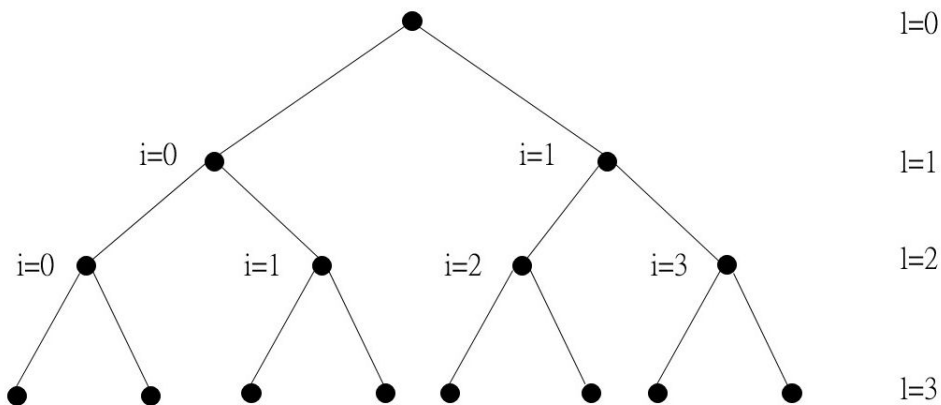
The decomposition splits original single freq bin into sub-bins, and merge time bins, as it can be seen in Figure 2.4, until the samples can no longer be divided by 2, that's why the maximum level of decomposition of 2^l samples is l .

The frequencies varies in (f_{min} and f_{max}), where f_{min} is zero and f_{max} is given by the Nyquist Frequency and the time interval is the same as the raw data interval, that is, how long the time series lasts.

The samples are arranged in 2^l layers, where:

- at maximum level l , the time resolution is given by $\Delta t_l = 2^l \Delta t_0$, and the frequency resolution is $\Delta f_l = 2^{-l} f_{max}$, so that means if $l = 0$ the resolution converges to the original signal representation,
- at the i^{th} layer, the node's frequency is in between the interval $i2^{-l} < \frac{f}{f_{max}} < (i + 1)2^{-l}$.

Figure 4.2 - This is an example of a signal wavelet decomposition of 3 levels. The original data is expressed in the first upper bin (level zero) and while the decompositions are being performed, the bins are divided into low and high frequency. All the bins can be identified with its respective decomposition level l and frequency layer i .



Decomposition must stop at level l if the number of samples is 2^l . In this dissertation, it is introduced a way to evaluate if going forward with the decompositions are advantageous or not for the representation in time-frequency of the signal.

The same way a signal can be decomposed in coefficients. These coefficients can be reconstructed in a time series through the Inverse Wavelet Transform. In the next section, these wavelet reconstructions are compared with the efficiency of a Fourier Transform reconstruction, and other applications are presented.

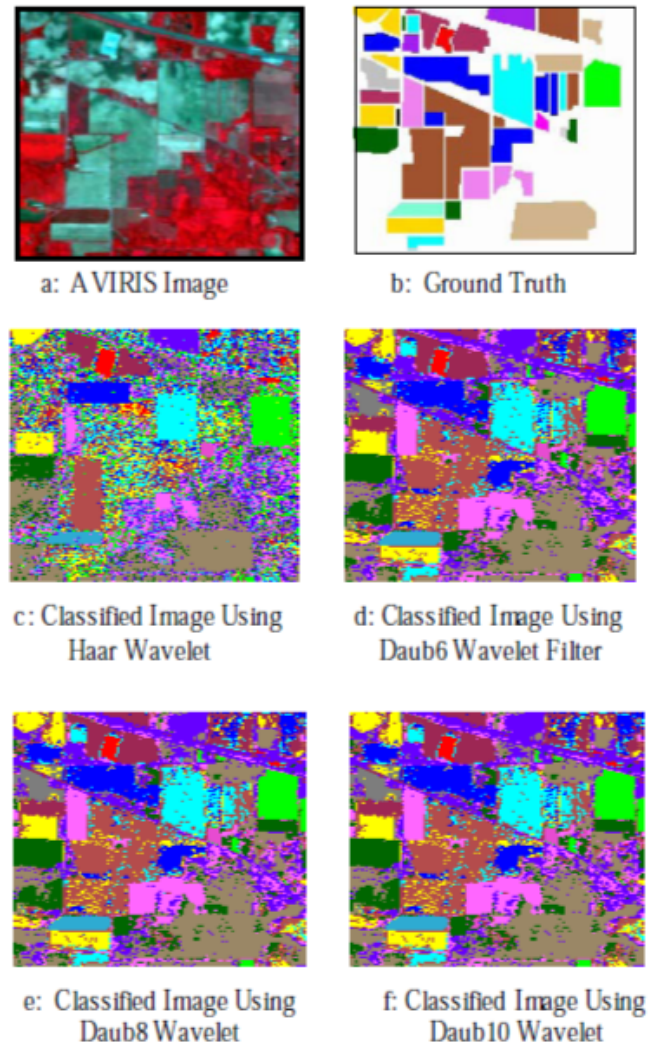
4.2 Applications

The Wavelet representation is extremely useful when it comes to signal processing, and the reason for that is with a few number of coefficients the signal can be represented really well. Further more, one can use the 3D Wavelet Transform to apply in images, as it is seen in Figure 4.3.

It is important to remember that if 100% of the wavelet coefficients are used, it must necessarily return to the original data, that goes for 2D or 3D. After all, there are no data losses in this type of representation.

Another interesting application is comparing the different mother wavelets efficiency used in the Wavelet Transform in images, as it is presented next (SHARIF; KHARE, 2014), where a image classification is made.














Figure 4.3 - AVIRIS image and the results for the wavelet analysis.



SOURCE: Sharif e Khare (2014).

In Figure 4.3, (a) shows an image produced by AVIRIS, an Airborne Visible and Infrared Imaging Spectrometer instrument (WANG et al., 2002) developed for Earth's remote sensing, and the objective of that scientific paper is to use wavelets in order to identify the ground features presented in Figure 4.4. The ground truth is given by (b) image, therefore the colors representing the data coefficients must be as close to (b) as possible. The Haar and Daubechie wavelets are used and the reconstruction plot is presented in (c), (d), (e) and (f).

Figure 4.4 - Ground truth colours.

Class	Colours used to represent	Producer Accuracy	User Accuracy
Stone/ Steel		94.5%	100.0%
Hay		98.4%	94.0%
Corn notill		67.91%	90.81%
Soya bean		83.5%	92.18%
Alfalfa		100.0%	100.0%
Wheat		21.5%	71.24%
Grass/ Pasture		100%	87.35%
Corn min		92.22%	46.26%
Soyabean- min		93.5%	92.15%
Woods		91.4%	98.0%
Soyabean-clean		92.8%	72.0%
Grass/ Trees		95.4%	94.94%
Bldg/ Grass		90.0%	87.42%

SOURCE: Sharif e Khare (2014).

The efficiency of all the mother wavelets used are listed in Figure 4.5, the classification accuracy between the different mother wavelets are very close, as the computational time.

Figure 4.5 - Comparison of Wavelet filters.

Table 1. Comparison of Wavelet Filters

Wavelet Filter	Computational Time(seconds)	Classification Accuracy(%)
Haar	45	83.8%
Db4	52	85.1%
Db6	54	85.3%
Db8	58	85.5%
Db10	60	85.6%

SOURCE: Sharif e Khare (2014).

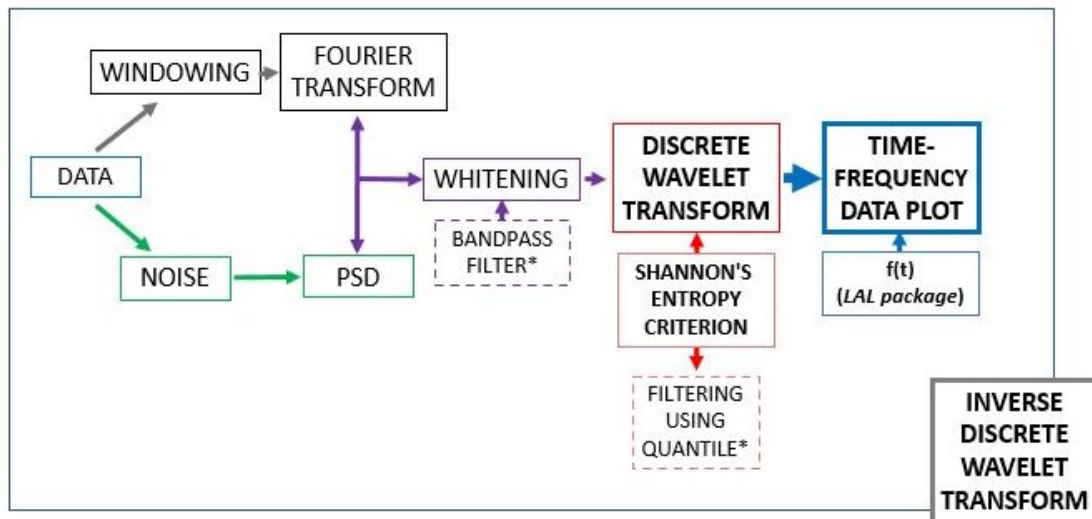
Applications of the Wavelet Transform was presented so far using images. In the next chapter the wavelet method will be applied for gravitational wave data. First, all the data analysis features will be used as tools for the signal treatment, and when the Discrete Wavelet Transform is performed, a decomposition stopping criterion called Shannon's Entropy will be used, creating a time-frequency representation with "tiles" and compared to the original LIGO data.

5 ADAPTIVE MULTIREOLUTION FOR WAVELET ANALYSIS

5.1 Methodology

In this work, a tool for wavelet analysis is developed and applied for gravitational wave events. The data analysis is performed through a code in *Python* programming language and it follows the steps of the procedure presented in Figure 5.1.

Figure 5.1 - Data analysis steps performed in this project starting from acquiring the data (input) and ending with the time-frequency data (output) plot. The dashed blocks represents optional steps that the user can choose to use it or not. The Inverse Discrete Wavelet Transform is executed for last in order to verify if the code's output returns to the input.



5.1.1 Data

The code steps defined in the above figure are applied for 10 gravitational wave events, which are listed in Table 5.1 with the parameters masses and signal-to-noise ratio (SNR).

Table 5.1 - The events used as input for the code and its features.

GW event	Mass 1 (M_{\odot})	Mass 2 (M_{\odot})	Signal to Noise Ratio
GW150914	35.6	30.6	24.4
GW170104	30.8	20.0	13.0
GW170817 (glitch)	1.46	1.27	33.0
GW190408_181802	24.6	18.4	14.7
GW190412	30.1	8.3	18.9
GW190519_153544	66.0	40.5	12.0
GW190521	95.3	69.0	14.4
GW190521_074359	42.2	32.8	24.4
GW190630_185205	35.1	23.7	15.6
GW190701_203306	53.9	40.8	11.6

The raw data is first downloaded in LIGO's Gravitational Wave Open Science Center, the website gives the user the option of choosing the gravitational wave event from two sampling rates (4kHz or 16kHz) and the data length in time (32s or 4096s).

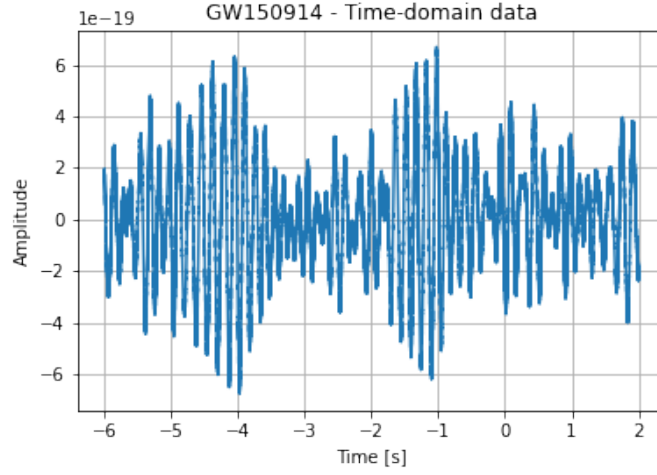
The downloaded data are samples that varies in time, that is, a time series. At this point, two chunks of data are chosen for different purposes:

- a) **Gravitational wave event:** samples from few seconds before and after the event that will be evaluated with and without a bandpass-filter and submitted to the discrete wavelet decomposition with a selection criterion and an adaptive filtering,
- b) **Noise:** samples at least 20 seconds before the event for short-duration signals (such as bursts) and 100 seconds for long-duration signals (such as continuous waves), so there are no gravitational wave information within. From this noise, the power spectral density is determined using 32 seconds of samples divided in 8 chunks of 4 seconds for the Fourier Transform.

Starting from **a)**, since the future steps in the analysis is to perform a Discrete Wavelet Transform, it is important to use a data chunk of time interval described as a number that is a power of 2. It is important to remember from the Gravitational Wave Theory chapter that the amplitude of the data is dimensionless.

In Figure 5.2, 6 seconds before the GW150914 event and 2 seconds after were chopped from the downloaded data and represented as a time-series. The amplitude varies as a power of 10^{-19} and it is clear that the event is immersed in noise.

Figure 5.2 - The GW150914 (H1) data expressed in time-domain, where $t = 0$ represents the moment the event occurs. The parameters used here are $rate = 4096\text{Hz}$, $N = 32768$ samples and $T = 8\text{s}$.



Before going to **b)**, the next step is to avoid *spectral leakage*, an appearance of spurious high frequency large amplitudes when performing a Fourier Transform using non-periodic signal on a finite interval (DIGITALCOMMONS, 2009), a matter to be discussed next.

5.1.2 Windowing

Given that the time-series was chopped from the raw data of 4096 seconds, the borders might experience spectral leakage and that problem can be fixed using the Windowing technique. In this code, the **Tukey window** was used for the gravitational wave search (ABBOTT et al., 2020).

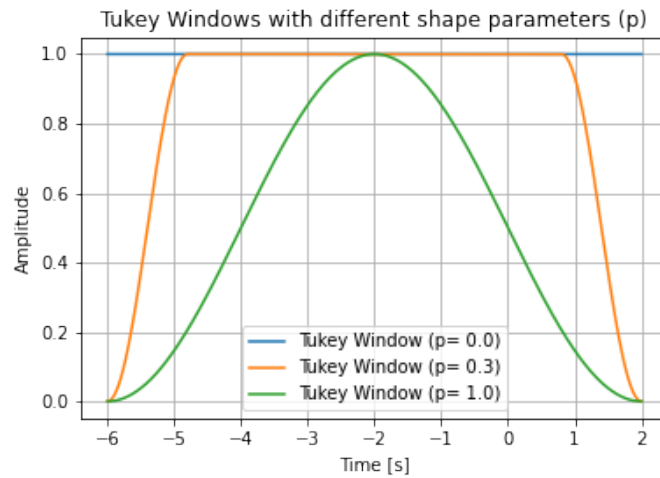
The Tukey window $w(x)$, also known as the cosine tapered window, can be understood as a cosine lobe of width $\frac{pN}{2}$ that is convolved with a Rectangular Window of width $N(1 - p/2)$.

$$w(x) = \begin{cases} \frac{1}{2} \left[1 - \cos \frac{2\pi n}{pN} \right], & 0 \leq n < \frac{pN}{2} \\ 1, & \frac{pN}{2} \leq n \leq \frac{N}{2} \\ w[N - n], & 0 \leq n \leq \frac{N}{2} \end{cases} .$$

The windowing consists in multiplying a function $f(t)$ with the value zero in the beginning and at the end of the window, eliminating the points that provokes the spectral leakage located in both beginning and end of the time-series.

While using the code, a shape parameter of the window must be given: p (as it is seen in the mathematical definition). It stands for the window's fraction inside the cosine tapered region. In the case of the Tukey Window, when this parameter p is zero, the window is equivalent to a Rectangular Window and if it's one, the window is equivalent to a Hann Window, a window function, presented in Figure 5.3.

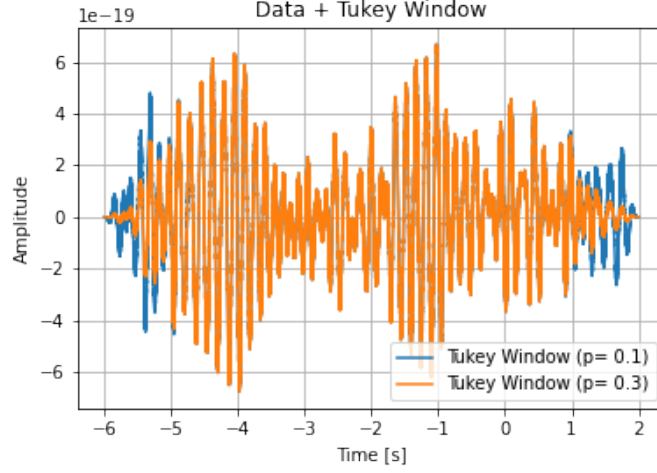
Figure 5.3 - The Tukey Window as a function of time, when analyzed with the shape parameters $p = 0$ (Rectangular Window), $p = 0.3$ and $p = 1$ (Hann Window).



If a high number for the shape parameter was considered, the window starts to influence the amplitudes near the gravitational wave event, as it is seen in Figure 5.4 that shows the result of the windowing procedure of the time-series. That is consequence of using the window, more specifically, when the shape parameter for the Tukey Window is 0.1, that means 0.5% of each end of the signal was affected (not set to 0 but reduced in amplitude) within the procedure, if it's 0.3, 15% of each end was affected and so on. This procedure modulates the input signal in order to the spectral leakage be evened out.

The amplitudes is expressed by a very small number, that can be a problem numerically and it is clear that the noise's amplitudes are as high as the signal's. But it can normalized using the Whitening technique, which will be seen next.

Figure 5.4 - The GW150914 (H1) windowed data expressed in time-domain when using a Tukey Window of shape parameters (p) 0.1 and 0.3.



5.1.3 Whitening

The whitening technique follows the procedure:

- Represent the time series containing signal and noise $s(t)$ in frequency-domain using the Fourier Transform: $\tilde{s}(f)$;
- Calculate the Power Spectral Density for the Noise: $S_n(f)$ from \mathbf{b} ;
- The whitened data in frequency domain ($\tilde{s}_w(f)$) is given performing the operation: $k \frac{\tilde{s}(f)}{(S_n(f))^{\frac{1}{2}}}$, where k is a constant given by Parseval's Theorem (IWASAKI, 2020),
- Since the next step is to use the DWT using a time series, it is necessary to perform the Inverse Fourier Transform in $\tilde{s}_w(f)$ and divide by dt (as a consequence of the Parseval's Theorem), resulting a whitened data in time-domain $s_w(t)$.

The whitening procedure is done to suppress the contribution to the time series from frequency bins which are more noisy. The noise power spectral density was generated by the code and presented in Figure 5.5 at a logarithmic scale.

When applying the whitening technique for the windowed GW150914 event (Figure 5.4, the result is expressed in Figure 5.6, where the noise's amplitudes becomes independent in frequency (just like a white noise) and in time-domain the noise contribution appears to be uniform.

Figure 5.5 - Power Spectral Density of the Noise as a function of the frequency.

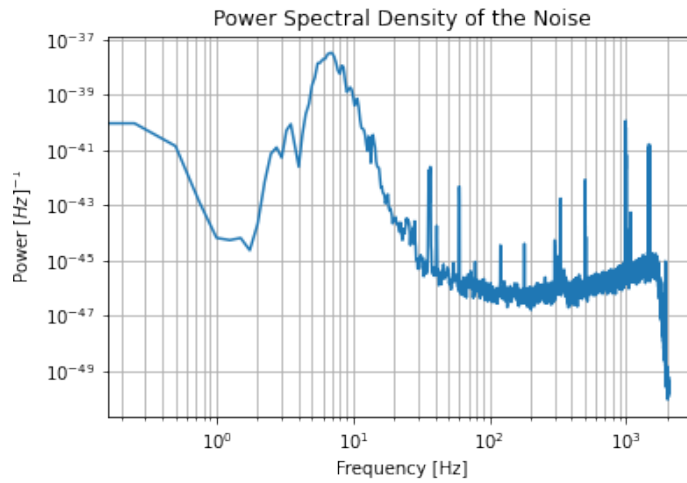
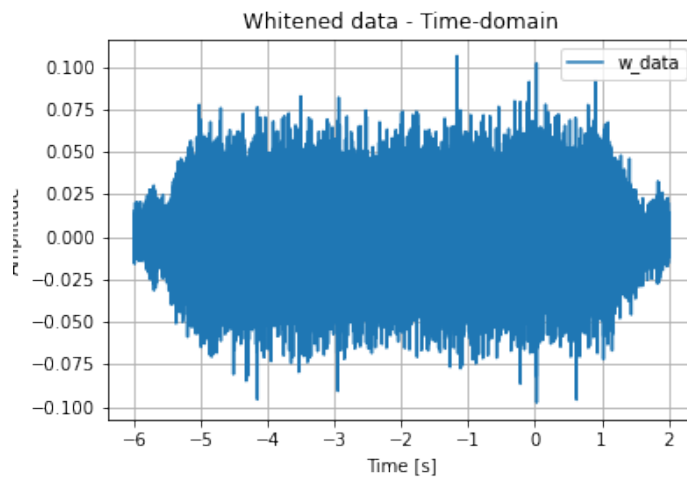


Figure 5.6 - The GW150914 (H1) windowed and whitened data expressed in time-domain.



The LIGO detector's frequency range is established between 10Hz to 10kHz ([ABERNATHY et al., 2018](#)). But there are a huge amount of instrumental noise between 10Hz and 30Hz. The events detected are located in a lower frequency band, more specifically smaller than 1kHz, therefore it is useful to erase the contribution of frequencies between those numbers. In order to perform this task, a bandpass filter is applied to the whitened data.

5.1.4 Bandpass filter

As the name says, a bandpass filter is a device of frequency filtering that is used in signal processing (ZHILYAKOV et al., 2020), where a range of frequency is established by the user and all the frequencies outside the bandwidth is rejected, that is, attenuated.

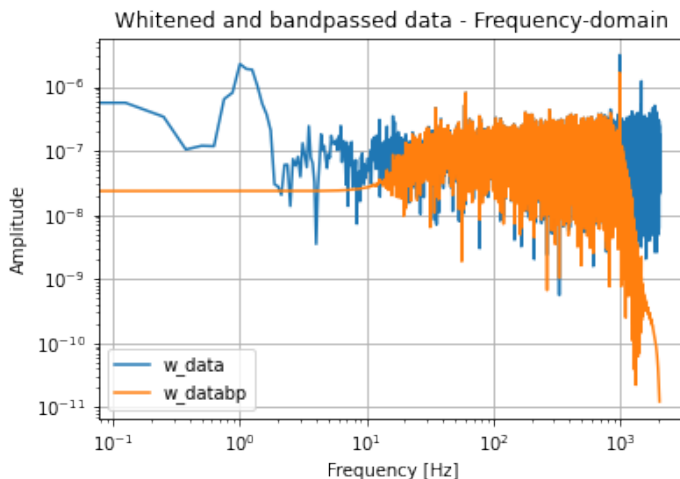
In this work, one bandpass filter was applied for the data:

- Filter 1: [20,1000] Hz,

Filter 1 can be executed for all the analyzed LIGO data, since the code wouldn't be eliminating any gravitational wave information using this bandpass.

A way of observing the effect of this filtering, a comparison between the whitened data with the whitened-bandpassed data of GW150914 in frequency domain is performed and it can be seen in Figure 5.7.

Figure 5.7 - The GW150914 (H1) event whitened and whitened-bandpassed (Filter 1) data comparison in frequency-domain.

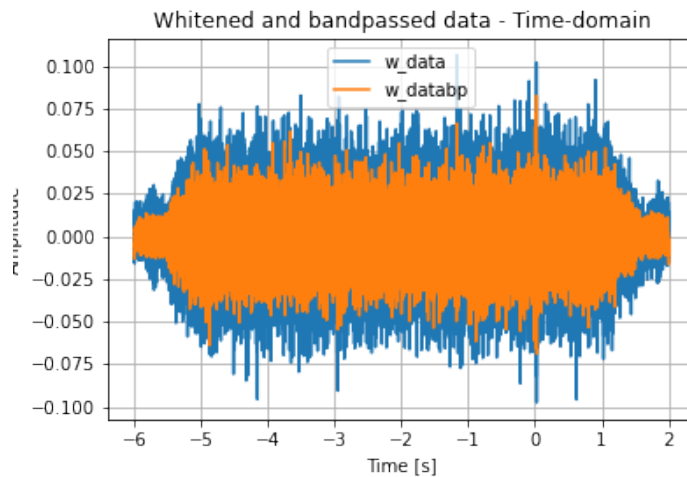


It is easy to identify the attenuated frequencies out of the Filter's 1 range. This will be reflected in the time-frequency plot and the differences between using filter and not using it will be discussed in those results.

A comparison in time-domain was also made, in order to verify the contribution in amplitude the deleted frequencies has. Figure 5.8 shows the results of both whitened

data and whitened-bandpassed data, the first one clearly has higher amplitudes related to higher and smaller frequencies.

Figure 5.8 - The GW150914 (H1) event whitened and whitened-bandpassed data expressed in time-domain.



Now that the data has been windowed, whitened and bandpassed, the basic data treatment is done, and the data is ready to go to the next step, which is the wavelet analysis.

5.2 Discrete wavelet transform

In this section, all the details of Chapter 4 that describes in a mathematical form the wavelet theory, will be used as a background for the numerical analysis.

Since the GW data are discrete, the discrete wavelet transform (DWT) is used in the code in Python. It's important to remember that the number of samples N to be analyzed are all a power of 2.

The signal is represented in a wavelet domain, which is localized in time and frequency. For every level and layer 3 parameters are generated:

- time coordinate: measured in seconds, it varies from $[0, T]$;
- frequency coordinate: measured in Hz, it varies from $[0, f_{NYQ}]$,
- coefficient: a vector of the same length as the input and a dimensionless parameter.

When the DWT is performed and the decompositions are being made, it divides the frequency bins and merges the time bins and Shannon's entropy criterion selects the maximum level of decompositions.

5.2.1 Shannon's entropy criterion

The form of Shannon's Entropy Criterion applied for a generic time-series is given by the following equation:

$$E^2(X) = - \sum_k \frac{x_k^2}{|X|^2} \log_2 \left(\frac{x_k^2}{|X|^2} \right), \quad (5.1)$$

where the entropy E is being calculated for a generic functional X is the norm.

In this work the entropy is used after the primary data processing (chopping, windowing and whitening) is performed. Since the next step is to apply the discrete wavelet transform in the data, while the code generates the decompositions, the entropy is calculated for the whitened data.

In the code, x_k represents the coefficients, $|X|^2 = \sum_i x_i^2$ is the norm of the coefficients, it is calculated from the sum of every two coefficients generated in the decomposition in every layer.

The maximum entropy is obtained as follows:

$$E(W_i^l) = - \sum \frac{1}{N} \log \left(\frac{1}{N} \right) = \log_2 N, \quad (5.2)$$

which means all coefficients are the same and it doesn't corresponds to a specific decomposition level.

The minimum entropy is given by

$$E(W_i^l) = - \log_2 1 = 0. \quad (5.3)$$

which corresponds to when all coefficients are zero except one and again, it doesn't match a specific decomposition level.

Since the tiles representation is used, the intensity of the signal is described with colours. More than that, an efficient signal representation using this criterion are: coefficients concentrated in the tiles generating a darker color in it, and with most

of the other coefficients approximately zero, that is close to white. The entropy will be the key point for the analysis that will generate this representation, defining if the decompositions will go forward or not.

The coefficients can be negative or positive, but all of them belong to the real axis.

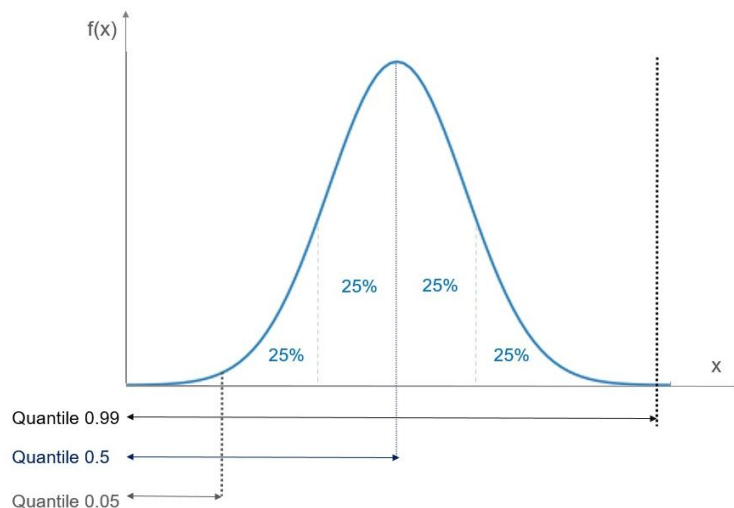
The entropy is calculated for $l = 0$, then $l = 1$, if the entropy of level 1 is higher than level zero, then the decomposition will stop and go back to level zero. This procedure will be repeated for every layer (black node) in each decomposition level of Figure 4.2 and its two "child" nodes at the following decomposition level. That leads to different shape of the tiles, and since their sides are measured with time and frequency resolutions, this representation will have a multiresolution that will adapt for every data used as input.

Furthermore, another filtering can be performed with the coefficients, in order to have a better multiresolution representation, this time using a tool called **quantile**.

5.2.2 Quantile threshold

In signal processing, quantiles are partitions in a distribution with equally probabilities that represents a certain percentage calculated from 0. to 1. In Figure 5.9 the quantile values are illustrated using a gaussian distribution.

Figure 5.9 - A quantile example using a simple gaussian distribution.



Mathematically, the quantile threshold is expressed as:

$$q = \int_{-\infty}^y |f(x)| dx \quad (5.4)$$

where $f(x)$ is a distribution and for a certain quantile q the point that represents this quantity in the distribution is given by y .

Therefore in the code, a fixed value for the quantile is defined and for each level, y is calculated and used as a reference for the filtering. All the values smaller than y in modulus for that level are filtered, that is, replaced by zeros. The quantile changes in every level, leading to an adaptive filtering of the coefficients with the quantile values within every layer.

Now that the code steps are defined, some of the successful representations are shown next.

5.3 Results: time-frequency representations

At this stage, the code was applied for the 10 of the strongest gravitational wave events and for each case, a list of coefficients was generated, with their corresponding time and frequency coordinates, as it was explained in the previous section.

In the time-frequency plot, the time coordinates are distributed in the x-axis, the frequency coordinates in the y-axis and the coefficients are represented as the colour code in the tiles and the colorbar next to every plot displays the colour with the lowest intensity represented by 0 and the highest represented by $\pm(0.2)$. All the plots are put in the same scale to identify which are the most and least intense emissions.

It is important to underline that due to the entropy's criterion applied in the code, a multiresolution in time and frequency is generated, being represented as tiles with different shapes, although respecting the area of $1/2$. With this flexibility, as the decompositions are being performed, it divides the frequency bins and merges the time bins and the entropy's criterion selects the maximum level of decompositions.

The actual gravitational wave chirp's signal is represented as a function of the frequency $f(t)$ and it is calculated from the equation below:

$$\dot{h}_+ h_\times - \dot{h}_\times h_+ = A^2 f \quad (5.5)$$

on the left side of the equation there are gravitational wave's polarization modes with its derivatives and on the right side $A = h_{\times}^2 + h_{+}^2$ represents the amplitude of the GW and f the frequency.

A simulation for the GW signal is made by using *LAL package* to calculate polarizations, using only the initial sources masses or the final event's mass and finally applying all the parameters in Equation 5.5, resulting in the frequency as a function of time $f(t)$.

From the 10 events, each of them was tested without a bandpass filter and with Filter 1. Table 5.2 displays the parameters used to generate the time-frequency representations.

In every time-frequency plot, the time axis are generated using the GPS corresponding to the time of the event as a reference, therefore the chirp instant corresponds to the mark zero in the axis.

The events are named after the date of observation, for example: GW150914 stands for "**G**ravitational **W**ave" and the detection date 2015-09-14). In some cases there are more numbers after the date of detection, such as: GW190701_203306, that means the detection occurred at the UTC time: 20h33'06". This notation is necessary in case other events are detected on the same day, for instance.

Table 5.2 - The code was applied for these 10 gravitational wave events and the parameters used to generate all the time-frequency representations are displayed in this table.

GW event	<i>Rate</i>	<i>T</i>	Tukey Window Parameter	Filters	<i>q</i>	Result
GW150914 (H1)	4096 Hz	2 s	0.2	No BP Filter 1	0.2	5.10
		8 s	0.3	No BP Filter 1	0.2	5.11
	16384 Hz	2 s	0.2	No BP Filter 1	0.2	A.1
GW170817 (L1) - glitch	4096 Hz	4 s	0.3	No BP	0.2	5.12
GW190521 (L1)	4096 Hz	2 s	0.15	No BP Filter 1	0.2	5.13
GW190521_074359 (L1)	4096 Hz	2 s	0.15	No BP Filter 1	0.2	5.14
GW170104 (H1)	4096 Hz	2 s	0.15	No BP Filter 1	0.2	5.15 5.16
GW190408_181802 (H1)	4096 Hz	8 s	0.3	No BP Filter 1	0.2	A.2 A.3
GW190519_153544 (L1)	4096 Hz	8 s	0.3	No BP Filter 1	0.2	A.4
GW190630_185205 (L1)	4096 Hz	8 s	0.3	No BP Filter 1	0.2	A.5 A.6
GW190412 (L1)	4096 Hz	8 s	0.3	No BP Filter 1	0.2	A.5 A.6
GW190701_203306 (L1)	4096 Hz	8s	0.3	Filter 1	0.2	A.9
		16 s	0.3	Filter 1	0.2	A.10

5.3.1 GW150914

This event was detected on 14 September 2015, as a consequence of a two binary black hole coalescence (ABBOTT et al., 2016), resulting in a black hole of $63.1 M_{sun}$ after the merger. It was the first detection of a GW event from the O1 and it was announced by the LIGO and Virgo collaborations on 11 February 2016. For that reason, the Physics Nobel prize was given to Kip Thorne, Barry C. Barish and Rainer Weiss for contributing in the development and success of the Laser Interferometer Gravitational-wave Observatory (LIGO) for the gravitational wave discovery.

In this project, since the resolutions of time and frequency are flexible and varies according to the input data, it is interesting to evaluate the event using different time intervals and sampling rates, after all, the minimum frequency resolution varies according to T^{-1} and the minimum time resolution is equal to $rate^{-1}$.

For this specific event, an evaluation using different time intervals (2s and 8s) and sampling rates (4kHz and 16kHz) was performed for the Hanford detector's data. The area of the tiles is still 1/2, but when using a higher sampling rate or a longer time interval, the maximum possible decomposition level is higher. As it will be seen next, these analysis won't significantly affect the time-frequency representations.

Figure 5.10 represents the event evaluated for 2s, 1 second before the event and 1 second after. The tiles with a more intense colour represents the high intensity of the data points. When analyzing the data without a bandpass filter, the high amplitudes corresponding to the noise in lower frequencies interferes in the time-frequency plot. On the other hand, when Filter 1 is used, the amplitudes in lower frequencies (below 20Hz and above 1000Hz) are suppressed, therefore, the noise is attenuated, leaving the signal information to be analyzed by Shannon's criterion and placed in the concentrated tiles.

When using a time interval of 8 seconds, 7 seconds before the event and 1 second after, the results are expressed in Figure 5.11, where it is clear that the intensity of the signal is very high, reaching a maximum tile of 0.20. Since T was increased, a tile with a smaller minimum frequency resolution than Figure 5.10 is expected and can be observed in both plots in lower frequencies. No modification were performed toward the time resolution, therefore the minimum distance between two consecutive time bins are the same for both Figures.

The left figure presents the signal with no filtering and because the signal-to-noise

ratio is smaller in 8s of data, the noise fluctuations appears in lower frequencies and is contained when Filter 1 is applied (right figure). In this filtered case, the noise intensity gets lower in smaller frequencies and the tiles adapt better to follow the signal's shape: horizontal in the inspiral and vertical in the merger.

Figure 5.10 - The results for GW150914 (H1) when evaluated using $T = 2s$ and a sampling rate of 4kHz. There are two time-frequency plots, the left figure representing the signal when no bandpass filter is applied and the right figure displays the event using the [20,1000]Hz filter. In both plots the event's $f(t)$ is presented as the blue curve.

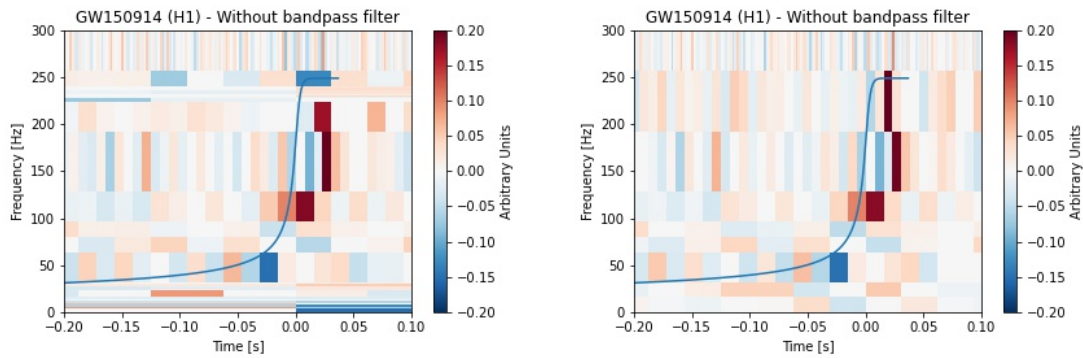
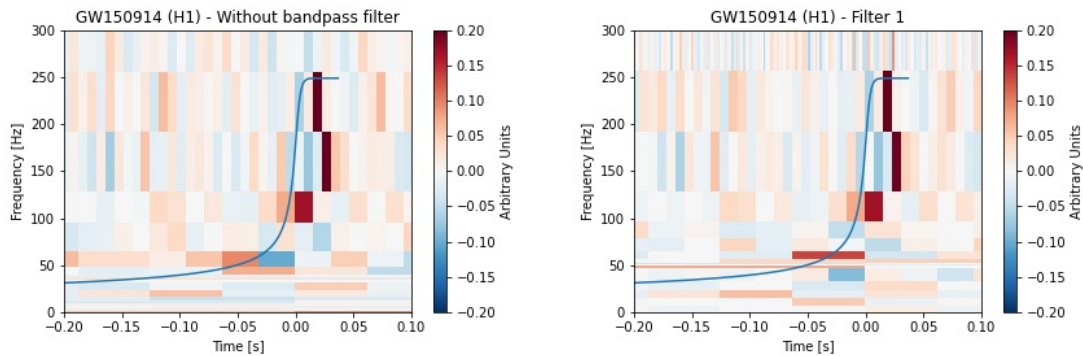


Figure 5.11 - The results for GW150914 (H1) when evaluated using $T = 8s$ and a sampling rate of 4kHz. There are two time-frequency plots, the left figure representing the signal when no bandpass filter is applied and the right figure displays the event using the [20,1000]Hz filter. In both plots the event's $f(t)$ is presented as the blue curve.



In both time intervals, a simulation for $f(t)$ was generated and it is present in all the plots as the blue curve. Even though noise fluctuations were detected when

analyzing the data, all the representations are very similar and they are coincident with the original chirp signal. Another point is that all representations are in a qualitative agreement with the spectrograms generated by LIGO. Those are factors that contribute for the idea that the developed code is accurate and provides solid time-frequency representations.

The analysis for the GW150914 event at Hanford using a sampling rate of 16kHz is presented in the first section of the attachment.

5.3.2 GW170817 (glitch)

The importance of the GW170817 (ABBOTT et al., 2017b) is far beyond than only being the first binary neutron star coalescence (NSNS) detected, this was the first event with an electromagnetic counterpart and the one with the longest duration (approximately 32 seconds). The merger of these two compact objects happened on 17 August 2017 and is most likely to result in a kilonova (KATHIRGAMARAJU et al., 2019) and this astronomical event is responsible for the generation of a short gamma-ray burst and also an extensive electromagnetic response.

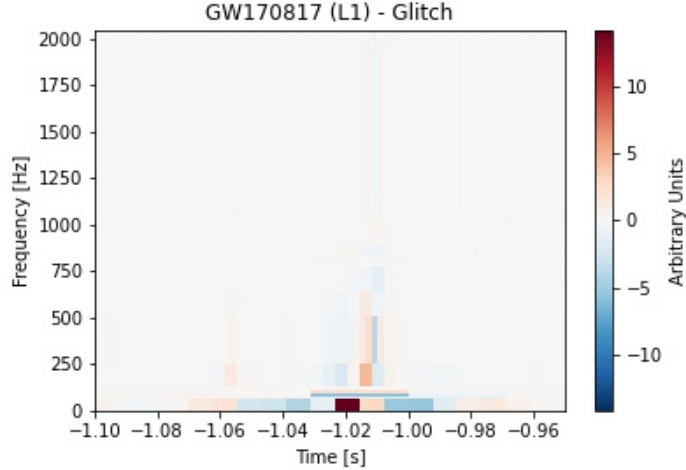
This event turned out to be a great breakthrough for multi-messenger astronomy, given the scientists' hopes of having to unite the gravitational waves and gamma-ray search.

When the interferometer in Livingston detected the event, an interesting fact happened, a **glitch** appeared seconds before the chirp. The chirp itself couldn't be identified when the code was applied for the L1 data, but the glitch representation is most accurate since it presents the highest SNR from the 10 events analyzed, as Figure 5.12 displays.

The glitch analysis was generated using the duration of it, which is 4 seconds, more specifically 3 seconds before the gravitational wave event and 1 second after.

In the figure, the colorbar scale is determined according to the highest coefficient of approximately 14, which appears at the lower center of the plot at -1.02 seconds. Using this scale it is clear that the background intensity is close to zero and there is a very intense presence in lower frequencies. This case is different than the other results, after all, it is representing not only the very low intensity gravitational wave emission that can hardly be identified in the plot, but as well as the glitch. The glitch presents a signal-noise-ratio of 33 in the Livingston detector, therefore it is expected that it would stand out in the time-frequency plot and cover the event.

Figure 5.12 - The GW170817 data with the glitch detected in the Livingston detector represented in time-frequency using $T = 4s$ and a sampling rate of 4kHz. The plot covers a time interval of $[-1.1, -0.95]s$ before the event, with 0 seconds being the gravitational wave emission. The glitch is centered in -1.02 seconds and the colorbar scale is defined from the maximum coefficient generated during the wavelet decomposition.



The glitch's position is very easy to identify, just like in the correspondent LIGO spectrogram (ABBOTT et al., 2017b), it is centered a second before the gravitational wave event, represented by -1s in the time-frequency plot.

The glitch extends itself from the minimum frequency (f_{min}) to the maximum (f_{NYQ}), that is $[0, 2048]Hz$ and possess the highest power in lower frequencies.

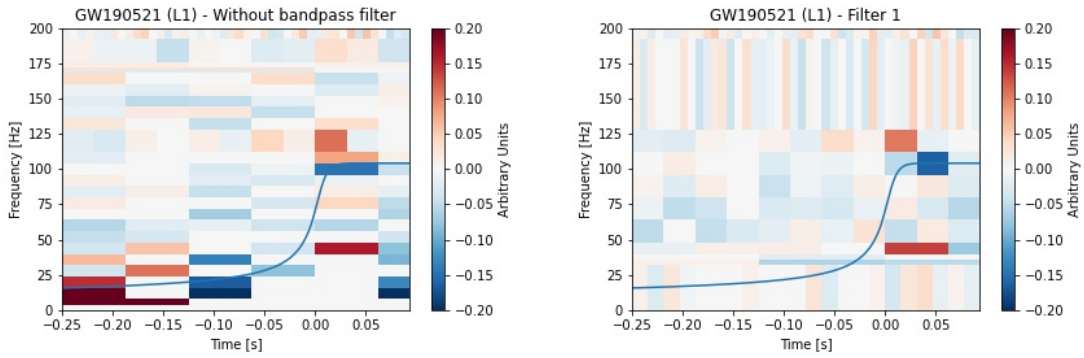
5.3.3 GW190521

This event marked another breakthrough in astrophysics for detecting the very first intermediate-mass black hole. Before this discovery, the final masses after the compact binary coalescences did not overcome $100M_{sun}$, but with masses of approximately $69 M_{sun}$ and $95.3 M_{sun}$, the largest progenitors observed so far, generated a black hole of $142M_{sun}$, providing 9 solar masses of energy radiated in the process in gravitational wave form (ABBOTT et al., 2020) on 21 May 2019.

This gravitational wave was analyzed by chopping 2 seconds of the Livingston detector's data, 1 second before the event and 1 after and since this event was very energetic and a signal with short-duration, the Tukey Window parameter was 0.15, so that it wouldn't eliminate the gravitational wave contribution.

Figure 5.13 presents the results for this analysis when no bandpass filter is applied and when Filter is used in the code. Both representations extends itself from frequencies lower than 25Hz up to 130Hz and the tiles containing GW signal are very clear and easy to identify.

Figure 5.13 - The results for the GW190521 event represented in time-frequency when using $T = 2$ s and a sampling rate of 4kHz. On the left, there is the time-frequency plot when no bandpass filter is applied and on the right, Filter 1 was used ([20,1000]Hz). In both plots the event's $f(t)$ is presented as the blue curve.



When no bandpass filter is used, the time-frequency plot maximum coefficient exceeds the module of 0.20, which are located in the both bottom left and right sides in frequency near zero as the darker tiles corresponding to noise contribution.

When Filter 1 is applied, the noise fluctuations decreases and since a short time interval of 2 seconds is used it is easy to see the clear improvement in the event's representation, filtering the lower frequencies, making the background stable and highlighting the tiles containing the GW event with a high intensity.

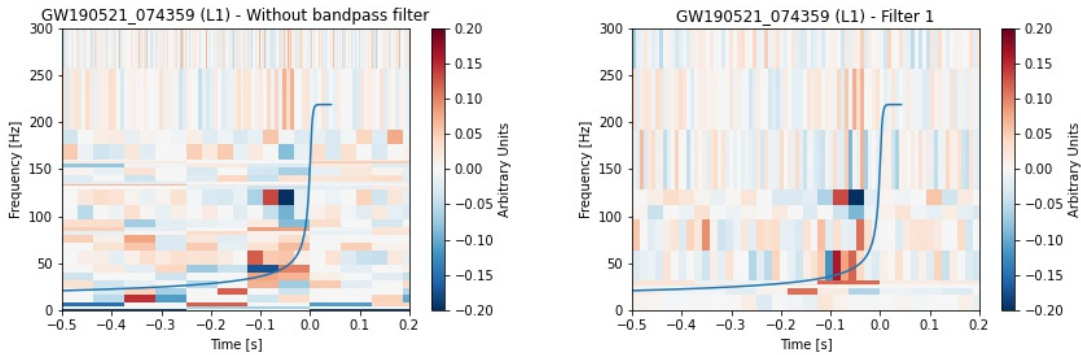
5.3.4 GW190521_074359

An event detected in the O3a run, GW190521_074359 (ABBOTT et al., 2021) was also generated by an inspiral and merger of a BBH source, leaving a black hole of final mass $71.1 M_{sun}$ also on 21 May 2019, more precisely at 07h43'59". Even though this signal is quite strong, the data contains a lot of noise, hence the noise intensity in the tiles are sometimes comparable with the signal's. When the filters are applied, the signal components in lower frequencies appears more concentrated in the tiles, which can be seen in Figure 5.14.

The analysis performed used a time interval of 2 seconds and a Tukey Window parameter of 0.2 and also filtering. The left figure presents noise coefficients in lower frequencies whose intensity is as higher as the gravitational wave's.

The noise amplitudes are can overcome the chirp's when using all frequency bands. On the other hand, when the filtering is applied, the amplitudes corresponding to frequencies from the $[20,1000]$ Hz interval are eliminated, therefore the right figure offers a much more clearer representation, where the signal is extended from approximately 25Hz up to almost 250Hz.

Figure 5.14 - Time-frequency representations for the GW190521_074359 event using $T = 2$ s and a sampling rate of 4kHz. On the left, there is the time-frequency plot when no bandpass filter is applied and on the right, Filter 1 was used ($[20,1000]$ Hz). In both plots the event's $f(t)$ is presented as the blue curve.



An important observation is that when using Filter 1, the tiles are vertically longer than the left Figure, that means the filter eliminated not only the frequencies outside the filtering interval, but also some of the noise fluctuation. While the decompositions were being made, no important contribution was found in the background corresponding to the chirp when the filter was applied, forcing the entropy criterion to focus on the GW contribution and providing the better tile shape for representing it.

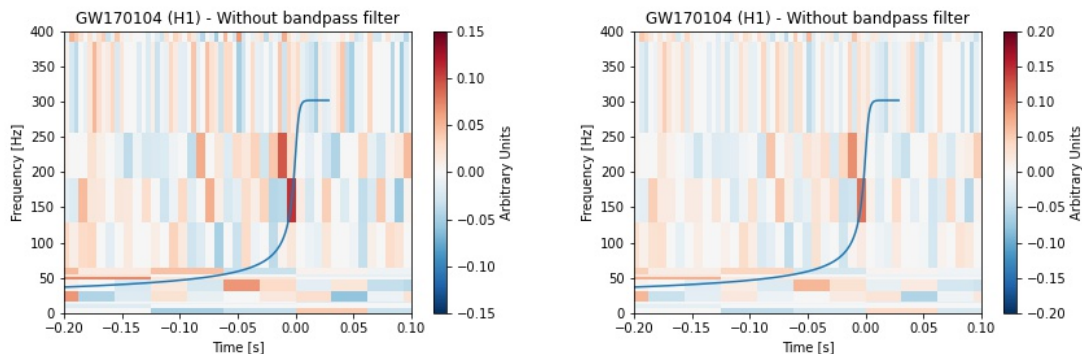
When comparing the LIGO version of this data (ABBOTT et al., 2021) presented as a spectrogram, the signal presents itself as vertically long, a property that the code seems to fulfil a better time-frequency plot than a signal horizontally long.

5.3.5 GW170104

This event was also detected on LIGO's first run, O1, more specifically the third event ever discovered. The event occurred due to an inspiral and merger of two black holes (BBH) of masses $30.9 M_{sun}$ and $20 M_{sun}$, resulting in a black hole of final mass $48.7 M_{sun}$, with two solar masses being radiated in gravitational wave form (ABBOTT et al., 2017a) on 4 January 2017.

Figure 5.15 presents the time-frequency representations for the GW170104 (H1) data when evaluated using a time interval of 2 seconds, 1 second before the event and 1 seconds after. The Tukey window shape parameter used in the data processing was 0.15 and no bandpass filter was used to generate those representations.

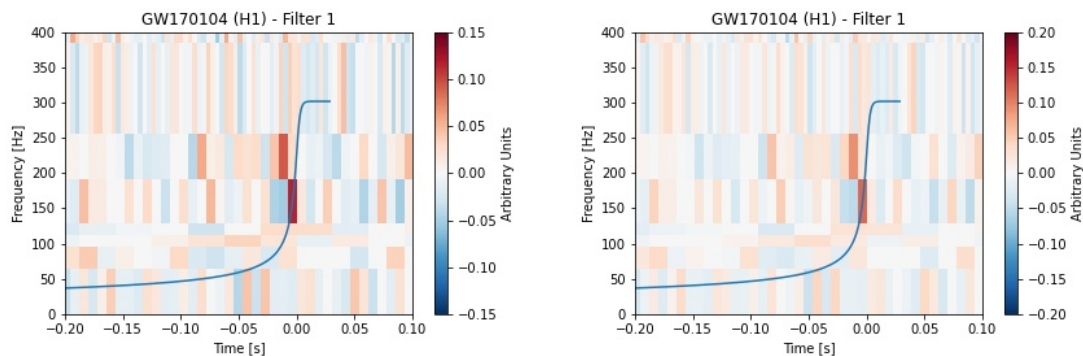
Figure 5.15 - Time-frequency representations for the GW170104 (H1) event using $T = 2s$, a sampling rate of 4kHz, with no bandpass filter and with different colorbar scales. The right figure will be compared with the other events in order to compare their intensities. In both plots the event's $f(t)$ is presented as the blue curve.



The chirp varies in approximately $[40,330]$ Hz, with a very concentrated tile which intensity reaches coefficients of modulus 0.11 between frequencies of $[130, 200]$ Hz. In order to compare the intensity of the events presented in this dissertation, the colorbar scale was also generated and presented on the right figure in Figure 5.15.

Figure 5.16 displays the results of the GW170104 event when using a Tukey window shape parameter of 0.2, and a bandpass filter that attenuates frequencies out of the $[20,1000]$ Hz range. One can observe that in lower frequencies the decompositions stopped in a lower level than Figure 5.15, after all, the tiles there are much more divided in smaller ones.

Figure 5.16 - Time-frequency representations for the GW170104 (H1) event using $T = 2$ s, a sampling rate of 4kHz and a bandpass filter of [20,1000]Hz and with different colorbar scales. The right figure will be compared with the other events in order to compare their intensities. In both plots the event's $f(t)$ is presented as the blue curve.



Since this specific chirp is long in frequency, that is, vertically long, the code represents it as a tile with a very intense colour that stands out in the analyzed time window in all plots.

Even when using different Tukey windows shape parameters, the code offers a very solid time-frequency plot, with small differences due to the noise fluctuations.

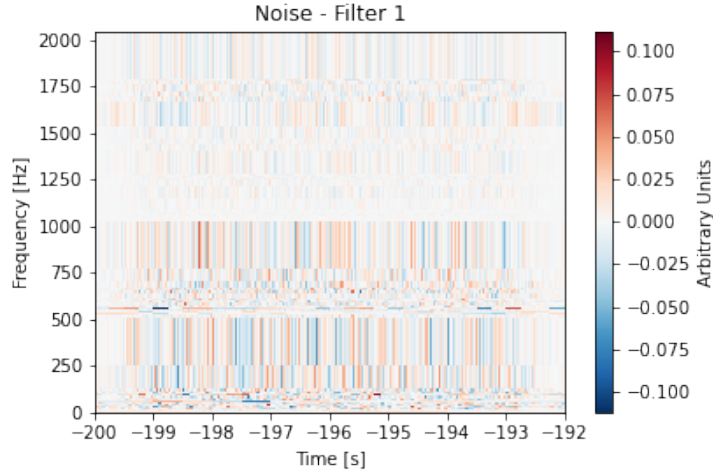
5.3.6 Noise

Instead of using a gravitational wave event as the code's input, the noise is also evaluated in time-frequency domain. Figure 5.17 displays the result of this representation when using the Filter 1.

The frequencies varies between [0,2048]Hz. Frequencies below 20Hz and above 1000Hz are attenuated as a consequence of Filter 1. That is the reason for the tiles in that region being lighter (lower intensity) than the ones between this interval.

It is easy to conclude that there are no evidence of a gravitational wave information. After all, the maximum tile intensity is approximately 0.1 and there are no chirp configuration of the tiles (cluster of tiles with high intensities) at any frequency band.

Figure 5.17 - Time-frequency representations for the noise 200s before the GW150914 event, using $T = 8\text{s}$, a sampling rate of 4kHz and a bandpass filter of $[20,1000]\text{Hz}$.



5.4 Inverse discrete wavelet transform

It was said previously that there are no losses when performing the Wavelet Transform in any time series, and to prove that is true, an addition to the code for the calculation of the Inverse Discrete Wavelet Transform (IDWT) was executed.

The code uses as input all the coefficients generated from the DWT without any filtering from the quantile and proceeds with the IDWT calculation. The output is stored in an array (*coeffs_rec*) and compared with the original whitened-bandpassed time series (*w_databp*), as Figure 5.18 presents. The numerical comparison was also made through a simple calculation:

$$\text{diff} = \sum (w_databp - coeffs_rec)^2 \approx 10^{-30} ,$$

which is basically numerical error.

Another attempt was made with the IDWT extension: calculating the difference between the original whitened-bandpassed data with the filtered data using quantile generated by the IDWT, as it is seen next in Figure 5.19. This time it is clear that the difference between the curves is not close to zero such as a numerical error anymore:

$$\text{diff} = \sum (w_databp - coeffsfilter_rec)^2 \approx 10^{-2} .$$

Figure 5.18 - Inverse Discrete Wavelet Transform for the GW150914 (H1) coefficients. The difference between both original and IDWT data is represented with the *diff* curve.

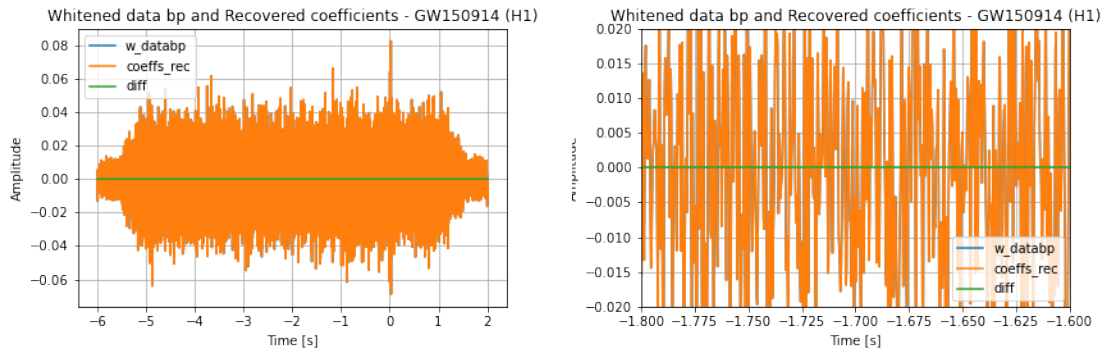
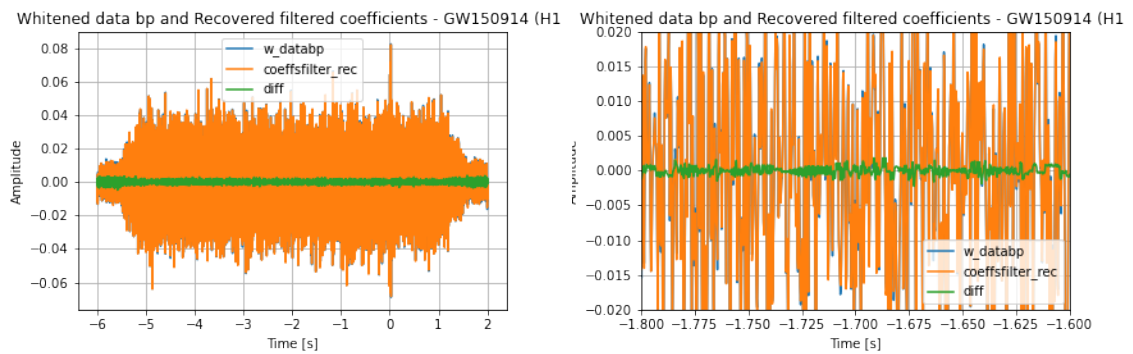


Figure 5.19 - Inverse Discrete Wavelet Transform for the GW150914 (H1) coefficients filtered with quantile and a bandpass filter ([20,1000]Hz). The difference between both original and IDWT filtered data is represented with the *diff* curve.



But if there was any doubt that the DWT code developed in this dissertation was working properly, this question is resolved when seeing the results in Figure 5.18 with an error of zero when comparing the input data (DWT) with the output data (IDWT).

5.5 Analysis

All the steps of the code development were shown in details in this chapter and the last procedure was to verify if the decompositions were being executed correctly, with the Inverse Discrete Wavelet Transform the algorithm was proven correct, but the accuracy in representing gravitational wave events must be analyzed.

5.5.1 Histogram

The analysis will be made from the histogram of the data's coefficients, after all, it contains all the information that generates the tiles in the time-frequency representations. This analysis was performed by generating a histogram for each of the 10 events presented during this dissertation, more specifically using the module of the coefficients. In order to avoid high coefficients corresponding to noise contribution, the bandpass filter data was used to generate the histograms.

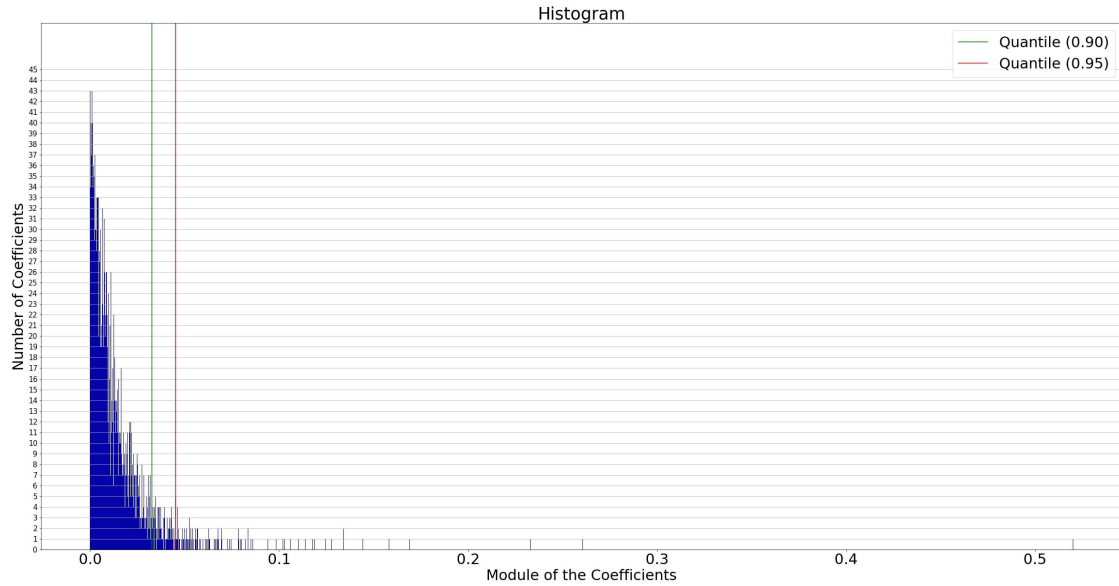
After the histogram was made, the events were divided in two categories:

- **GOOD REPRESENTATION:** There are a clean separation between the distribution of the background and foreground. In the histogram the event's coefficients can be easily identified as the highest or one of the highest, in case the data contains glitches or tiles with a high intensity of noise contribution, in general.

For example, when applying this procedure to GW150914, more specifically for the coefficients that generated the Filter 1 result in Figure 5.11, the result is displayed in Figure 5.20.

When the representation is considered good, then it is possible to identify the signal and eventually the noise contribution in the histogram, as follows in Figure 5.21.

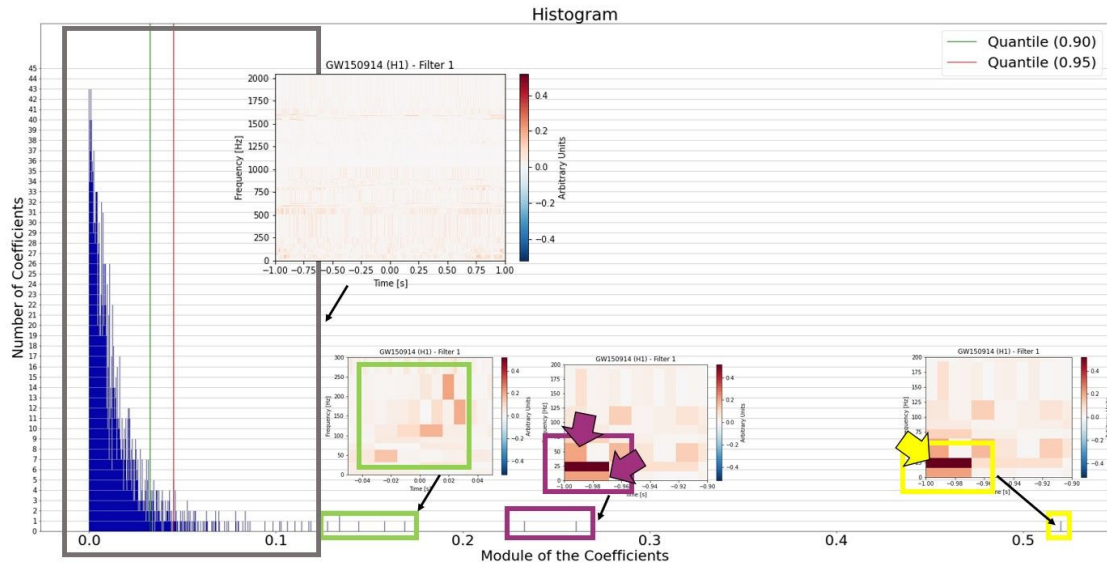
Figure 5.20 - Histogram of the GW150914 (H1) event's coefficients module. The quantile representing 0.90 and 0.95 of the coefficients curve is described as the green and red lines, respectively.



Since the noise contributions were very easy to identify and presented as a small quantity, the event's coefficients were next to have the highest intensity, therefore it was easy to analyze the histogram and jump into those conclusions.

This procedure was applied for the other 9 events and the ones that followed the same histogram pattern as GW150914 were: GW170104, GW170817, GW190519_153544, GW190521 and GW190521_074359, therefore they are considered a **good representation** that the code developed in this dissertation generated.

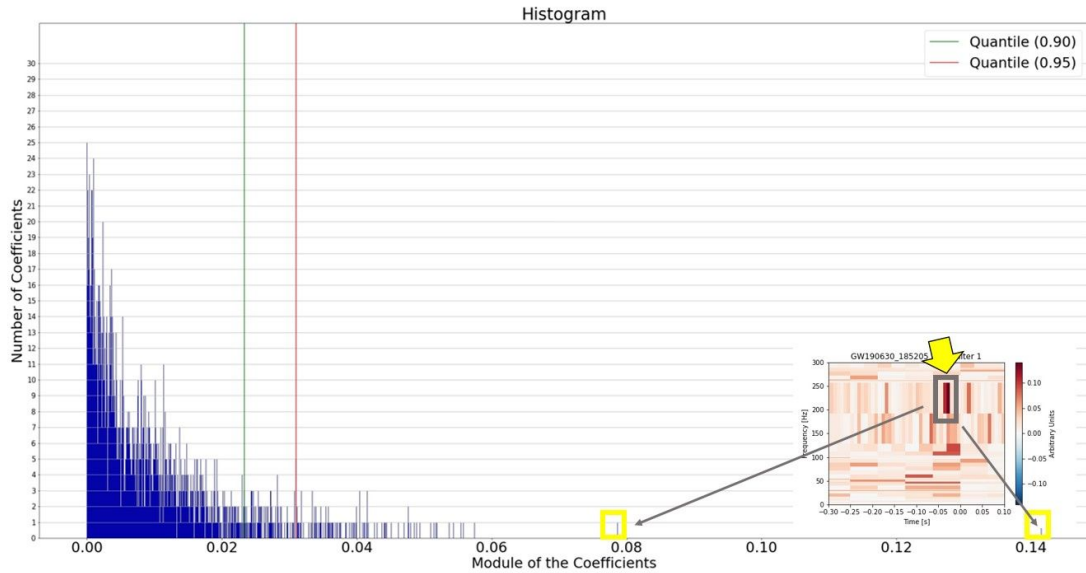
Figure 5.21 - Histogram of the GW150914 (H1) data's coefficients module. The coloured boxes indicates the correspondent coefficient in the histogram to the the time-frequency representation tile. The bigger time-frequency plot represents all the coefficients below $|0.13|$ and its correspondent in the histogram. The quantile representing 0.90 and 0.95 of the coefficients curve is described as the green and red lines, respectively.



- **UNSEPARATED REPRESENTATION:** Unlike the good representation, in this case the coefficients histogram presents itself as difficult to identify the event's contribution due to the intensity of the noise and the gravitational wave's are very similar, just like the GW190630_185205 event as the histogram shows in Figure 5.22.

It is clear that the noise and the signal's contribution are very close in module, with a smaller number of low coefficients, therefore trying to identify the signal's contribution is difficult. But, since the user can still identify at least two significant coefficients that corresponds to the event's tiles in the histogram, the adaptive multiresolution representation for this event is considered **unseparated**. Once again the same procedure was performed for all the events, the other 3 events that presented Figure 5.22's histogram pattern were: GW1900408_18102, GW190412 and GW190701_203306, considered then unseparated representations.

Figure 5.22 - Histogram of the GW190630_185205 (H1) event's coefficients module. The quantile representing 0.90 and 0.95 of the coefficients curve is described as the green and red lines, respectively. The highest coefficient is the only one that can be identified and it is presented with the yellow box.



The question now is: are there any specific conditions which makes the code more accurate and generates a good representation? In order to answer this, the parameters in Table 5.1 were used to perform three analysis:

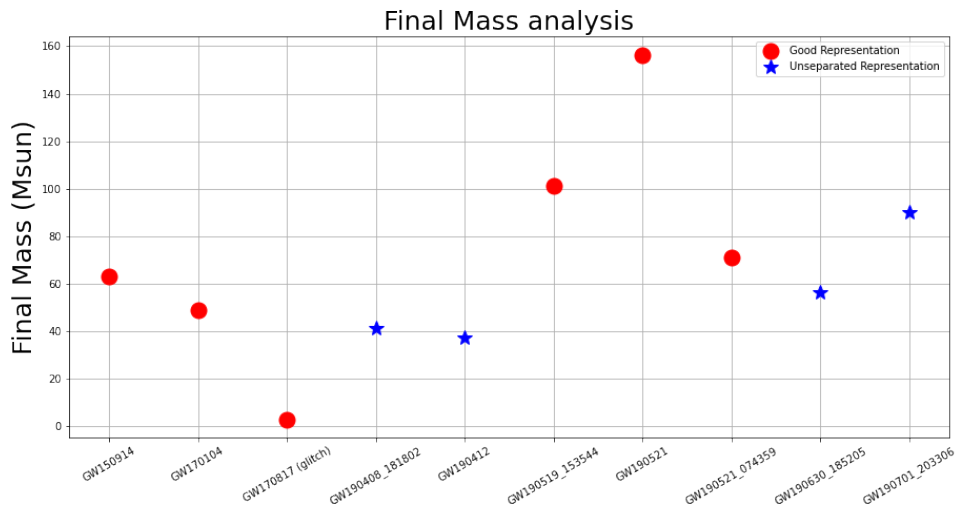
- **Final Mass:** the mass from the merger are put in the y-axis with its respective gravitational wave event in the x-axis;
- **Signal-to-Noise Ratio:** the network SNR is displayed in the y-axis with its corresponded gravitational wave event in the x-axis;
- **Mass-SNR:** the final masses are distributed in x-axis with its respective signal-to-noise ratio from the evaluated gravitational wave events in the y-axis,

the goal is to find an association with the gravitational wave parameters that when used as input for the code, the output provides the most accurate time-frequency representations.

5.5.2 Final mass

For this first item, when analyzing Figure 5.23, no pattern is observed, that means the code works for whichever final mass parameter, after all, a compact binary inspiral source generating a compact object (for those cases, a black hole) with final mass around $30M_{sun}$ and an object with final mass around $100M_{sun}$ both obtains a good representation when those data are processed with the developed algorithm.

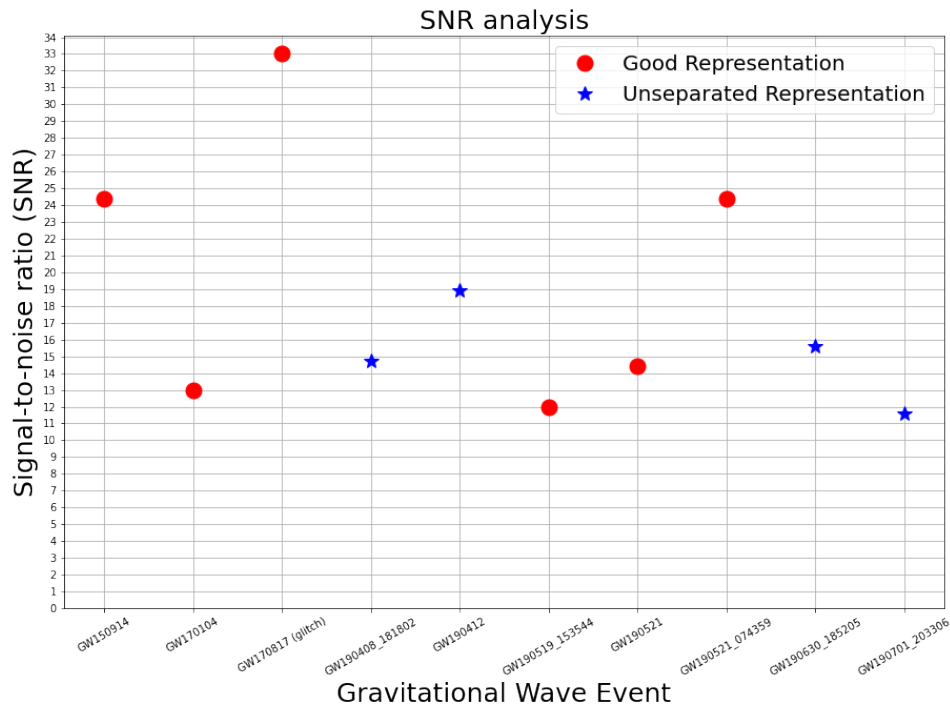
Figure 5.23 - Final mass analysis.



5.5.3 Signal-to-Noise Ratio

As for this second item, the signal-to-noise ratio is evaluated in Figure 5.24 and it is clear that if the event is correlated with a SNR bigger than 20, the code provides a good representation. This can be observed in the presented results, if the signal is very intense, as the GW170817 glitch, the algorithm provides a multiresolution time-frequency representation with a very concentrated signal in the tiles, as it was previously shown.

Figure 5.24 - SNR analysis.



5.5.4 Mass-SNR

As for the third and last item, the colours in Figure 5.25 stands for the evaluated events. Even though a concentration of the red points indicating the good representation was not found, that means no pattern between the network's SNR and the final masses of the events could be identified, the code only has a SNR limitation and it is not related to the final masses.

One last attempt is to use the mass ratio for the analysis, but even then no pattern was detected as it is observed in Figure 5.26.

Figure 5.25 - Mass-SNR analysis.

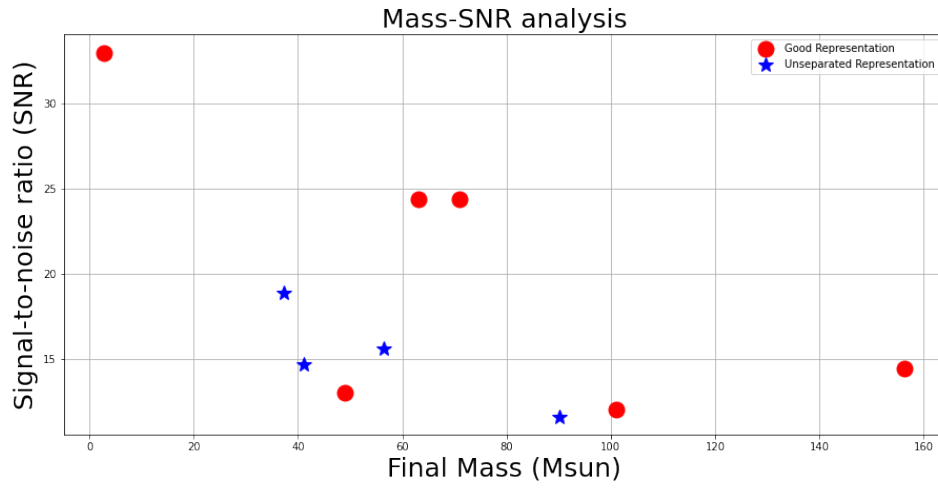
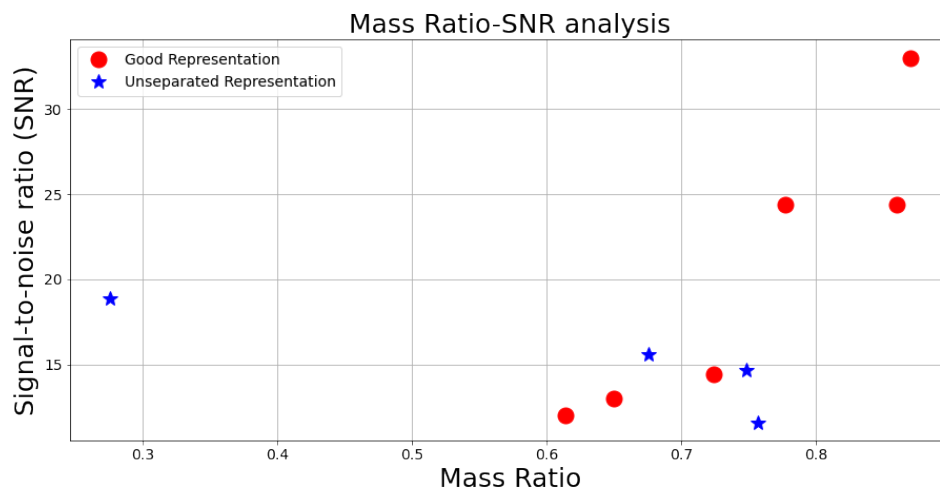


Figure 5.26 - Mass Ratio-SNR analysis.



6 CONCLUSION

This dissertation had the main objective of developing a code in Python to perform a wavelet analysis using the Shannon's Entropy as a selection criterion to represent any time-series in time-frequency domain. The developed code can be accessed in the following *Git-Hub* repository: <https://github.com/ana-guimaraes/Adaptive-Multiresolution-for-Wavelet-Analysis>.

This work first explored the theory of gravitational waves, starting from the very first prediction through Einstein's General Relativity Theory, giving the paths towards the equation that describes a gravitational wave and its physical consequences such as types of sources that generates different emissions and it answered the question of how an interferometer with a Michelson-Morley configuration with Fabry-Perot cavities such as LIGO and VIRGO detects it.

Then, the most important part of this work is the data analysis part. It is known that a signal can be described in time-domain and frequency-domain, but the resolution in both cases provides losses in frequency and time resolutions, respectively. This work suggests a compromise: an analysis using a time-frequency domain with the help of wavelets. An important advantage of using wavelets is that when the Wavelet Transform is executed for a time-series, it transforms every data point directly into the time-frequency plane, providing the time-frequency resolution plot using Heisenberg Boxes.

Since the wavelet domain represents a signal using tiles (or the technical term Heisenberg Boxes), with shape (width and height) given by time resolution and frequency resolution, respectively, the association between these two parameters is given by the Heisenberg-Gabor Uncertainty Principle, which establishes the area of this tile as $1/2$ for the wavelet domain. Using this fixed tile area, different time and frequency resolutions can be used for representing a signal, and a selection criterion for the decomposition is provided by the Shannon's entropy for the decompositions. The results can be represented in multiresolution plots that uses this resolution flexibility to adapt itself for whichever time-series is used as input and concentrates the signal at maximum in the lowest number of tiles.

A code in Python was developed following the presented ideas, starting from the initial data processing: acquiring the data, calculating the PSD for the noise, chopping, windowing, whitening and bandpassing (an optional step) the data and finally performing a wavelet analysis. Then, the Discrete Wavelet Transform was used with

the Shannon's Entropy as a selection criterion for the decompositions. The output of this procedure are: time coordinates, frequency coordinates and coefficients (a dimensionless parameter).

The three output information are put in a time-frequency plot, with the coefficients representing by colour code. An additional optional step is to filter coefficients below a quantile threshold, in order to eliminate the coefficients with relatively small intensity. And to verify if the procedure was developed correctly, the Inverse Discrete Wavelet Transform was executed using the time and frequency bins and the coefficients as input, in order to resume to the original time-series.

This code was applied for 10 of the 52 gravitational (the additional 2 are the two BH-NS published last week) detected by the LSC so far, chosen from the highest SNRs. The time-frequency representations were sorted out in 2 categories: good and unseparated representations, depending how effective the time-frequency representation is to separate the background and foreground.

The results for the time-frequency representation were satisfactory, the events with the highest intensities, GW150914, GW190521 and the glitch from GW170817 presented a very accurate result, with well concentrated tiles representing the chirp, making it easier to the user to identify in the data a gravitational wave information.

Specifically for the first event GW150914 (H1) that generated the most satisfactory result, two attempts were performed: changing the time resolution using different sampling rates (4kHz and 16kHz) and changing the time intervals (2 and 8 seconds) for the frequency resolution. In all cases the code provided very similar results and as a consequence of that, even changing the parameters, such as the time interval that reduces the minimum frequency resolution generating a smaller tile in height (frequency) and when changing the sampling rate, it reduces the minimum time resolution and generates a smaller tile in width (time), the code still offers a solid representation, which only validates this very interesting and useful tool.

Even for the least intense signal considered, GW190701_203306, the representation was not able to separate satisfactorily foreground and background but showed well the noise contribution grouped in frequencies below 70 Hz around the event.

It was also clear that the filtering in all cases helped eliminate the noise contribution (as in lower frequencies and noise fluctuations), even though it is an *ad hoc* step, it contributed attenuating the amplitudes of the frequencies outside the [20,1000]Hz

range and provided a better time-frequency plot rather than without using it.

Also in all analyzed cases, the tiles representing the chirp was overlapped with the instantaneous evolution of frequency of the emitted GW according to the waveform best fit ($f(t)$) used as a visual reference.

The time-frequency representation were considered as good and not well separated, depending on how effective the wavelet decomposition was in separating background and foreground. It is clear that the good representations offers a histogram with most of the coefficients close to zero, leaving the highest intensities to the event and to occasional glitches that appears in the plot, as it is seen in the GW150914, GW170104, GW170817, GW190519_153544, GW190521 and GW190521_074359 events.

An unseparated representation also offers high intensity coefficients that can be identified as the event's, but most of the coefficients are drowned in the noise contribution, making it difficult to identify it all, as it is shown in the GW190630_185205, GW1900408_18102, GW190412 and GW190701_203306 events.

After analyzing all the time-frequency representation, it is clear that the code works at best when performing an analysis of a short-duration signal, with high intensity (SNR) and the code represents signals better when they are stretched vertically, i.e. have a rapidly varying frequency.

A final analysis was made by trying to correlate the gravitational wave's parameters: final mass, mass ratio with the SNR, with the code's accuracy. Within all cases, the code shows no correlation with the parameters, which means it is useful and works with all GW's informations.

In every events Shannon's selection criterion was essential for the representation's refinement, concentrating the signal and the noise using the lowest number of tiles, making it easier to identify their contributions. It is important to notice that the developed tool is not only useful in the gravitational wave representation but also for other types of data, after all, all it needs is a time-series as input.

A further step would be to create a clustering criterion for high intensity coefficients, to identify candidate events.

REFERENCES

- ABBOTT, B. et al. Ligo: the laser interferometer gravitational-wave observatory. **Reports on Progress in Physics**, v. 72, n. 7, 2009. 1, 15
- ABBOTT, B. P. et al. Observation of gravitational waves from a binary black hole merger. **Physical Review Letters**, v. 116, 2016. ISSN 10797114. 1, 13, 14, 16, 52
- _____. Gw170104: observation of a 50-solar-mass binary black hole coalescence at redshift 0.2. **Physical Review Letters**, v. 118, 2017. ISSN 10797114. 58
- _____. Gw170817: observation of gravitational waves from a binary neutron star inspiral. **Physical Review Letters**, v. 119, 2017. ISSN 10797114. 16, 54, 55
- _____. Gwtc-1: a gravitational-wave transient catalog of compact binary mergers observed by ligo and virgo during the first and second observing runs. **Physical Review X**, v. 9, 2019. ISSN 21603308. 16
- _____. A guide to ligo-virgo detector noise and extraction of transient gravitational-wave signals. **Classical and Quantum Gravity**, v. 37, 2020. ISSN 13616382. 22, 29, 41
- ABBOTT, R. et al. Gw190521: a binary black hole merger with a total mass of 150 m. **Physical Review Letters**, v. 125, 2020. ISSN 10797114. 55
- _____. Gwtc-2: compact binary coalescences observed by ligo and virgo during the first half of the third observing run. **Physical Review X**, v. 11, 2021. 16, 56, 57, 81, 82, 84
- ABERNATHY, M. R.; LIU, X.; METCALF, T. H. An overview of research into low internal friction optical coatings by the gravitational wave detection community. **Materials Research**, v. 21, 2018. ISSN 15161439. 28, 44
- BISHOP, N. T.; REZZOLLA, L. Extraction of gravitational waves in numerical relativity. 2016. Disponível em:
<<http://dx.doi.org/10.1007/s41114-016-0001-9>>. 11
- BRADASCHIA, C. et al. The virgo project: a wide band antenna for gravitational wave detection. **Nuclear Instruments and Methods in Physics Research, A**, v. 289, 1990. ISSN 01689002. 16, 18
- BUTKOV, E.; SPOSITO, G. **Mathematical physics**. [S.l.: s.n.], 1969. ISSN 0031-9228. 23, 33

- CAI, R. G.; CAO, Z.; GUO, Z. K.; WANG, S. J.; YANG, T. The gravitational-wave physics. **National Science Review**, v. 4, 2017. ISSN 2053714X. 10
- CHO, A. Remembering joseph weber, the controversial pioneer of gravitational waves. **Science**, 2016. ISSN 0036-8075. 1
- DIGITALCOMMONS. **The discrete Fourier transform, part 4: spectral leakage**. 2009. 23-34 p. Disponível em: <<https://digitalcommons.fairfield.edu/engineering-facultypubs>>. 41
- D'INVERNO, R. A.; HARVEY, A. **Introducing Einstein's relativity**. [S.l.: s.n.], 1993. ISSN 0031-9228. 5
- DIRKES, A. Gravitational waves - a review on the theoretical foundations of gravitational radiation. **International Journal of Modern Physics A**, v. 33, 2018. ISSN 0217751X. 6
- EINSTEIN, A. Relativitätsprinzip und die aus demselben gezogenen folgerungen. **Jahrbuch der Radioaktivität**, v. 4, 1907. 5
- FIELD, S. E.; GALLEY, C. R.; HERRMANN, F.; HESTHAVEN, J. S.; OCHSNER, E.; TIGLIO, M. Reduced basis catalogs for gravitational wave templates. **Physical Review Letters**, v. 106, 2011. ISSN 00319007. 12
- GRIFFITHS, D. **Introduction eletrodynamics**. [S.l.: s.n.], 2010. ISSN 1079-7114. 7, 10
- HEINBOCKEL, J. H. **Introduction to tensor calculus and continuum mechanics - preface**. [S.l.: s.n.], 1996. ISSN 00218936. 5
- HU, X.-C.; LI, X.-H.; WANG, Y.; FENG, W.-F.; ZHOU, M.-Y.; HU, Y.-M.; HU, S.-C.; MEI, J.-W.; SHAO, C.-G. Fundamentals of the orbit and response for tianqin. 2018. Disponível em: <<http://dx.doi.org/10.1088/1361-6382/aab52f>>. 18
- IWASAKI, A. Deriving the variance of the discrete fourier transform test using parseval's theorem. **IEEE Transactions on Information Theory**, v. 66, 2020. ISSN 15579654. 43
- JOFFRE, E.; WEALTHY, D.; FERNANDEZ, I.; TRENKEL, C.; VOIGT, P.; ZIEGLER, T.; MARTENS, W. Lisa: heliocentric formation design for the laser

interferometer space antenna mission. **Advances in Space Research**, 2020. ISSN 18791948. 18

KATHIRGAMARAJU, A.; GIANNIOS, D.; BENIAMINI, P. Observable features of gw170817 kilonova afterglow. **Monthly Notices of the Royal Astronomical Society**, v. 487, 2019. ISSN 13652966. 54

KEMAO, Q. Two-dimensional windowed fourier transform for fringe pattern analysis: principles, applications and implementations. **Optics and Lasers in Engineering**, v. 45, 2007. ISSN 01438166. 31

LANGANGEN; VASKINN, A.; SKAGERSTAM, B. S. Interference of light in a michelson-morley interferometer: a quantum optical approach. **International Journal of Optics**, v. 2012, 2012. ISSN 16879384. 17

LÜCK, H. The geo600 project. **Classical and Quantum Gravity**, v. 14, 1997. ISSN 02649381. 18

MACLENNAN, B. Gabor representations of spatiotemporal visual images 2 the gabor uncertainty principle. **Science**, 1994. 25

MAGGIORE, M. **Gravitational waves. volume 1: theory and experiments**. [S.l.: s.n.], 2009. ISSN 0001-7701. 10, 17, 21, 26

MALLAT, S. **A wavelet tour of signal processing**. [S.l.: s.n.], 2009. 2, 31, 32, 33

MATVEEVICH, M. P. Wavelet analysis statistical data. **Advances in Sciences and Humanities**, v. 1, p. 30–44, 2015. 34

NUSSENZVEIG, M. **Curso de física básica 4 - ótica, relatividade, física quântica**. [S.l.: s.n.], 2014. 7

OLIVEIRA, N. F.; AGUIAR, O. D. The mario schenberg gravitational wave antenna. **Brazilian Journal of Physics**, v. 46, 2016. ISSN 16784448. 15

PUNTURO, M. et al. The third generation of gravitational wave observatories and their science reach. **Classical and Quantum Gravity**, v. 27, 2010. ISSN 13616382. 18

ROTHAN, H. A.; BYRAREDDY, S. N. The epidemiology and pathogenesis of coronavirus disease (covid-19) outbreak. **Journal of Autoimmunity**, v. 109, 2020. ISSN 10959157. 16

SHARIF, I.; KHARE, S. Comparative analysis of haar and daubechies wavelet for hyper spectral image classification. **International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives**, XL-8, 2014. ISSN 16821750. 35, 36, 37

STURANI, R.; TERENCEZI, R. Adaptive multiresolution for wavelet analysis. **Journal of Physics: Conference Series**, v. 122, 2007. 2

THE role of gravitation in physics. **Physics Today**, v. 10, 1957. ISSN 0031-9228. 1

WANG, P.; LIU, K. Y.; CWIK, T.; GREEN, R. Modtran on supercomputers and parallel computers. **Parallel Computing**, v. 28, 2002. ISSN 01678191. 36

YADAV, A. **Nyquist-Shannon sampling theorem**. [S.l.: s.n.], 2009. 24

ZHILYAKOV, E. G.; BELOV, S. P.; OLEINIK, I. I.; BABARINOV, S. L.; TRUBITSYNA, D. I. Generalized sub band analysis and signal synthesis. **Bulletin of Electrical Engineering and Informatics**, v. 9, 2020. ISSN 23029285. 45

APPENDIX A

A.1 GW150914 - 16kHz

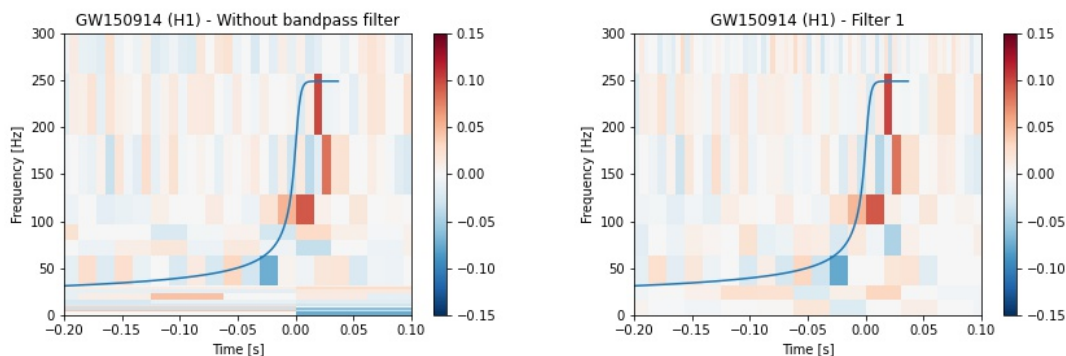
In Chapter 5, the GW150914 event was evaluated for a sampling rate of 4kHz and different time intervals, so that the minimum frequency resolution can be tempered with. Another attempt is to temper with the minimum time resolution, and since $\Delta t_{min} = 1/rate$, an analysis with a sampling rate of 16kHz is performed.

In order to generate the time-frequency resolution, the developed code was applied for the 16kHz GW150914 (H1) data, using 2 seconds of data, 1 second before the event and 1 second after. The result is presented in Figure ??, where the signal was analyzed without and with a bandpass filter.

As it was expected, the noise in frequencies below the bandpass filter range [20,1000]Hz are attenuated, the intensity of the tiles representing the chirp reaches almost a module of 0.15 and just like the 4kHz samples, it is coincident with the original chirp signal in time-frequency plane, $f(t)$, calculated with the *LAL package*.

The signal respects the same frequency limits of approximately [30,270]Hz and for this event, the minimum time resolution wasn't very useful, perhaps with a more vertical event, this sampling rate could be more efficient.

Figure A.1 - Time-frequency representations for the GW150914 event using 2 seconds and a sampling rate of 16kHz. The left figure shows the time-frequency plot without filtering and the right figure presents the result for the wavelet representation when using a [20,1000]Hz bandpass filter. In both plots the event's $f(t)$ is presented as the blue curve.



The following events were all detected by the LIGO's O3a run.

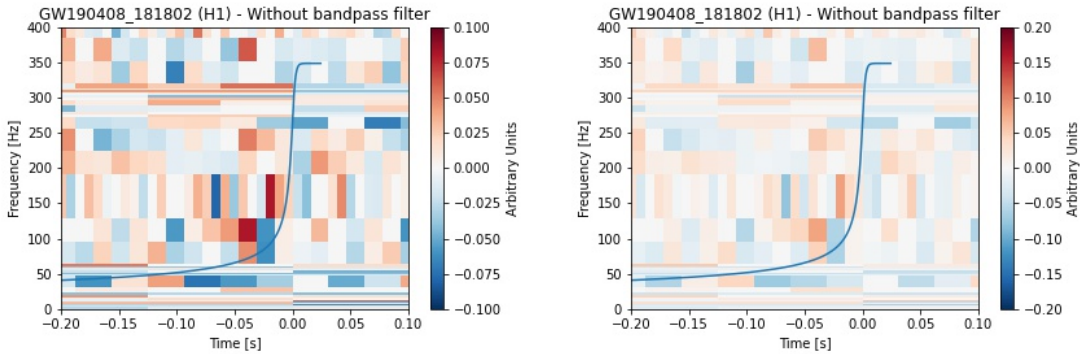
A.2 GW190408_181802

As it was previously said, this event was one of the 39 events detected in the O3a run. The gravitational wave was produced by the inspiral and merger of two black holes of masses $35 M_{sun}$ and $18 M_{sun}$, generating a black hole of mass $53.9 M_{sun}$ and about 2 solar masses in gravitational wave form on 8 April 2019 at 18h18'02".

When downloading the data from LIGO's website, it was observed that the event is a bit long horizontally when compared to the others, and in order to make the tiles smaller in frequency, that is, to define a minimum frequency resolution, a time interval of 8 seconds was chosen to fulfil this task.

The results for this event evaluated with a time interval of 8 seconds, 7 seconds before and 1 second after the event and without any bandpass filter is displayed in Figure A.2, where the data contains a lot of noise, but it is still possible to identify tiles with a higher intensity corresponding to the chirp, as it is seen near the zero mark in time and frequencies between (100,200)Hz. The tiles with the concentrated intensity along the signal in time-frequency, $f(t)$ (presented as the blue curve) corresponds to the noise that the event presents when captured by the Hanford's detector.

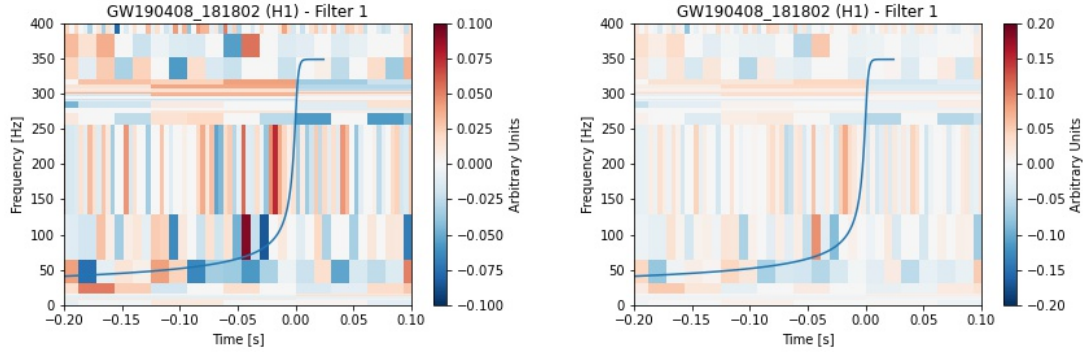
Figure A.2 - Time-frequency representation of the GW190408_181802 (H1) data using $T = 8s$, a Tukey Window shape parameter 0.1, no bandpass filter and a sampling rate of 4kHz. The left figure presents the maximum coefficients corresponding to $|0.8|$, but since a comparison between the events intensity is made, a plot with the colorbar of the other events is shown in the right figure. In both plots the event's $f(t)$ is presented as the blue curve.



Due to noise fluctuations, when the bandpass filter is applied in this event, the time-frequency representation, as it is shown in Figure A.3, shows a more intense and concentrated tile in frequencies (90,110)Hz, an effect of the Shannon's criterion

for the filtered data, since the noise was eliminated with the bandpass filter. The noise in the background now is represented as tiles (vertically) long in frequency, attenuated the noise contribution and made the more intense tiles corresponding to the event easier to identify.

Figure A.3 - Time-frequency representation of the GW190408_181802 (H1) data using $T = 8\text{s}$, a Tukey Window shape parameter 0.1, a bandpass filter ($[20,1000]\text{Hz}$) and a sampling rate of 4kHz. The left figure presents the maximum coefficients corresponding to $|0.8|$, but since a comparison between the events intensity is made, a plot with the colorbar of the other events is shown in the right figure. In both plots the event's $f(t)$ is presented as the blue curve.



Once again, the time-frequency representations on the right side of Figures A.2 and A.3 has the colorbar with intensity of module 0.2 in order to be compared with the intensity of the other 9 events.

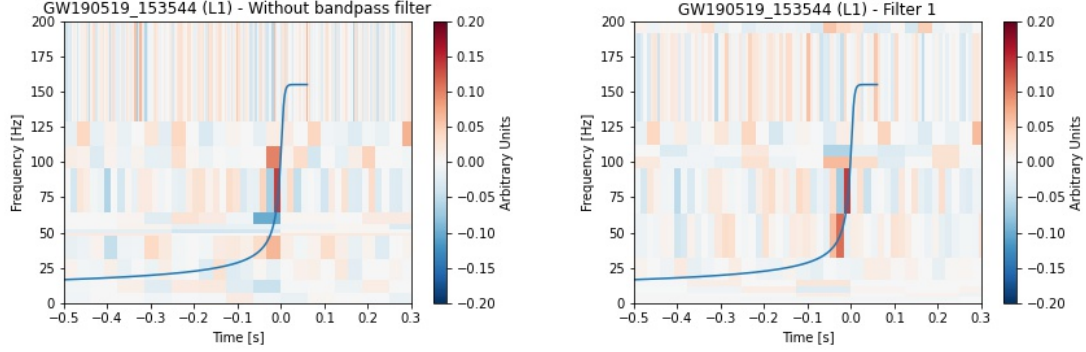
A.3 GW190519_153544

Another one of the 39 events from the detection list of the O3a run, this gravitational wave event was generated by the inspiral and merger of a BBH source, which produced a black hole of final mass $106.6 M_{sun}$ on 19 May 2019 at 15h35'44".

The event was evaluated for the two conditions: without bandpass filter and Filter 1 $[20,1000]\text{Hz}$ and the results are presented in Figure A.4. It is clear the importance of the filter for this case, it assists in the refinement of the signal in lower frequencies and also decreases the noise contribution due to fluctuations in all frequency bands.

In both representations the maximum coefficient has a module of 0.16 and when the filter is applied, it clear that the entropy criterion is essential to the much more refined multiresolution plot.

Figure A.4 - The time-frequency representations for the GW190519_153544 (L1) event, when using $T = 8\text{s}$, a Tukey window shape parameter of 0.3 and a sampling rate of 4kHz. On the left figure there is the result for this event evaluated without a bandpass filter and on the right the result for the event when Filter 1 ($[20,1000]\text{Hz}$) is applied in the data processing. In both plots the event's $f(t)$ is presented as the blue curve.



A.4 GW190630_185205

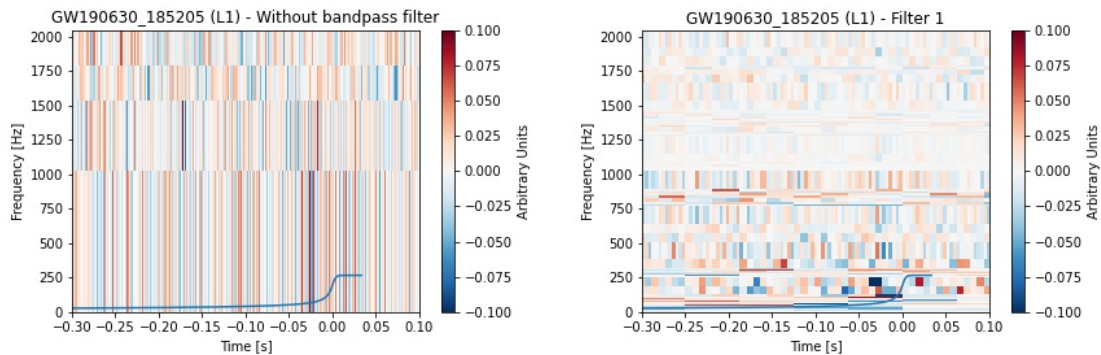
Another one of the events detected in the O3a run, this gravitational wave event was generated by the inspiral and merger of a binary black hole of masses $35.1 M_{sun}$ and $23.7 M_{sun}$ generating a black hole of mass $59.1 M_{sun}$ on 30 June 2019 at 18h52'05".

To generate the results for this event, considering that the inspiral is also horizontally long, the best way to pick a time interval T is to take that into consideration in order to generate a small minimum frequency resolution, but if T is too big, the signal-to-ratio of the data will be very small, since too much noise will be taken into account. For that reason, the results were generated using 8 seconds, 6 seconds before the event and 2 after.

The time-frequency representations for this event are presented in A.2 and A.3, where a deadlock is reached: when analysing the data without a bandpass filter, the decompositions stopped at level 3 as A.5 (left) shows and the maximum coefficient was 0.1 exactly near the mark zero in time. That means the entropy criterion chose that time-frequency plot as the best representation due to the data parameters used. A way to change this result is to switch the parameters such as Tukey window, use another time interval and sampling rate, or applying a filter in order to attenuate the amplitudes in certain frequencies. On the right side of A.5 is the result when applying a bandpass filter as it was suggested.

A filter that attenuates frequencies out of the [20,1000]Hz range is applied and as a consequence of that, the tiles corresponding to the event becomes concentrated around $t = 0$ s. If that time-frequency plot was generated without the user knowing it had a gravitational wave event inside, that configuration would raise the question if an event was producing those concentrated and intense tiles.

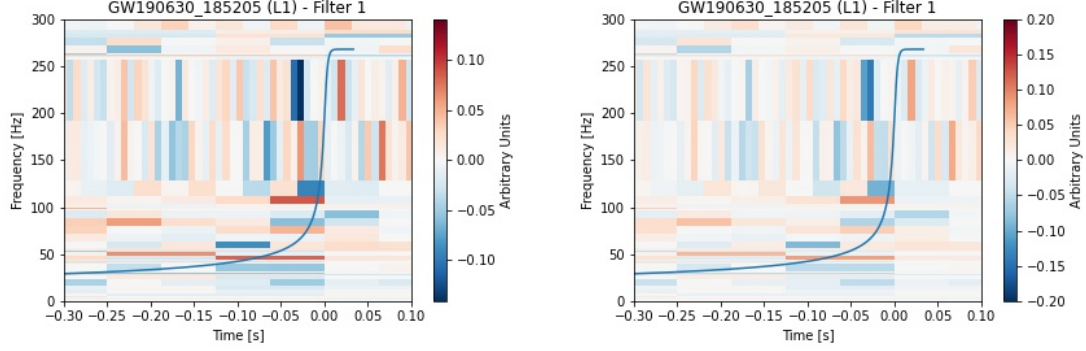
Figure A.5 - The time-frequency representations for the GW190630_185205 (L1) event, when using $T = 8$ s, no bandpass filter a sampling rate of 4kHz. On the left figure there is the result for this event evaluated without a bandpass filter and on the right the result for the event when Filter 1 ([20,1000]Hz) is applied in the data processing. In both plots the event's $f(t)$ is presented as the blue curve.



When generating a window in time [-0.3,0.1]s and frequency [0,300]Hz to see the event a bit closer, the result is presented in Figure A.6, where the tiles with a higher intensity stands out representing the event.

Since this event is horizontally long as it is seen in the LIGO's Collaboration paper (ABBOTT et al., 2021), the code offers the best representation using Shannon's Entropy selection criterion, but it is clear that the code provides a most accurate plot when the chirp is vertically long. Even so, it still offers a solid representation.

Figure A.6 - The time-frequency representations for the GW190630_185205 (L1) event, when using $T = 8\text{s}$, a bandpass filter ($[20,1000]\text{Hz}$), a sampling rate of 4kHz and different colorbar scales. In both plots the event's $f(t)$ is presented as the blue curve.



A.5 GW190412

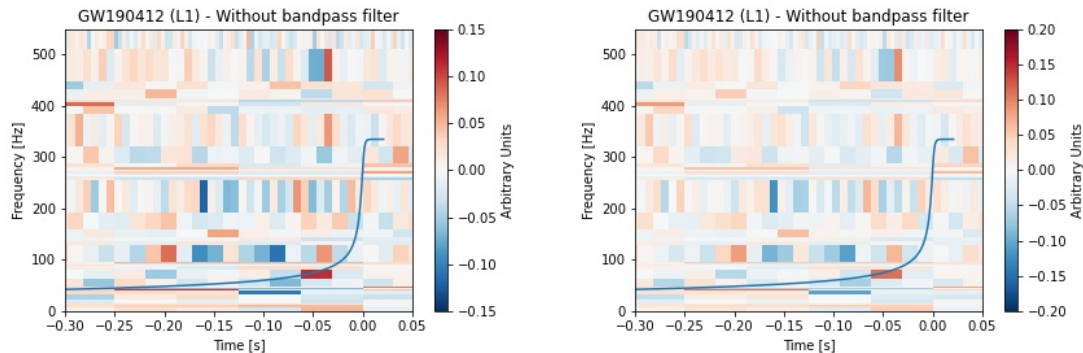
This event was also detected in the O3a run (ABBOTT et al., 2021), it was the first gravitational wave event generated by an inspiral and merger of a binary black hole with very different sizes, more specifically approximately $30.1 M_{sun}$ and $8.3 M_{sun}$. This event happened on 12 April 2019.

In order to generate a time-frequency representation with the developed code, a time interval of 4 seconds was used, 3 seconds before the event and 1 after the data from Livingston. The selection criterion was again very useful, specifically for this case, where just like GW190408_181802 there are a lot of noise in certain frequencies in the raw data.

Figure A.7 shows the results for the time-frequency representation for the event when no bandpass filter is used, it is clear that there are a lot of noises and the amplitudes of all frequency bands weren't attenuated, therefore the noise fluctuation made it difficult to identify the chirp's coefficients. Even then, when comparing to the background, the user can presume there might be a gravitational wave signal around the $[30,400]\text{Hz}$ frequency range.

When applying a bandpass filter that attenuates frequencies outside the $[20,1000]\text{Hz}$ range, the final result was very refined, the selection criterion given by Shannon's entropy made the time-frequency plot to select the tiles where the intensity was highest, preventing the background to overcome the signal, providing the time-frequency plot in Figure A.8. Lower frequencies tiles are much more attenuated, concentrating

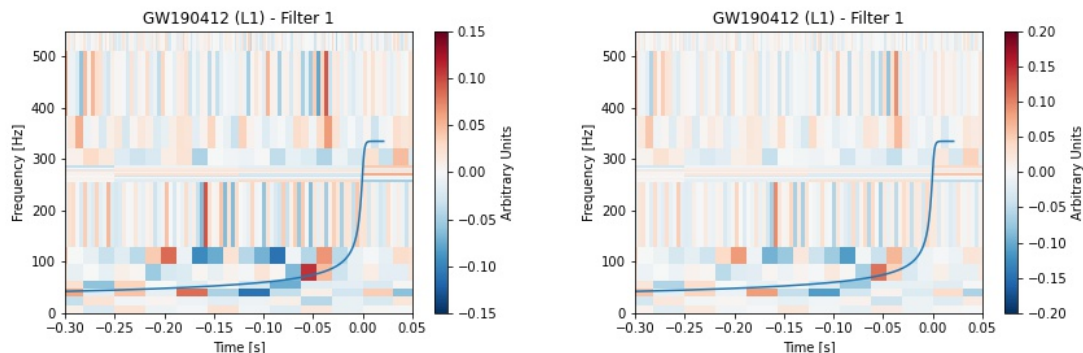
Figure A.7 - The time-frequency representations for the GW190412 (L1) event, when using $T = 4s$; without a bandpass filter; a Tukey window shape parameter of 0.3, a sampling rate of 4kHz and different colorbar scales. In both plots the event's $f(t)$ is presented as the blue curve.



the tiles horizontally along the $f(t)$ (blue curve).

The glitches that appears around $(-0.20, 0.15)s$ and $(150, 300)Hz$ is better represented when the bandpass filter is applied, where it is concentrated in a long tile in frequency. The same happens with the glitch around $(-0.05, 0)s$ and $(350, 520)Hz$ represented as tiles with a more intense colour and with a position better determined in Figure A.8.

Figure A.8 - The time-frequency representations for the GW190412 (L1) event, when using $T = 4s$, a bandpass filter $([20, 1000])Hz$, a sampling rate of 4kHz and different colorbar scales. In both plots the event's $f(t)$ is presented as the blue curve.



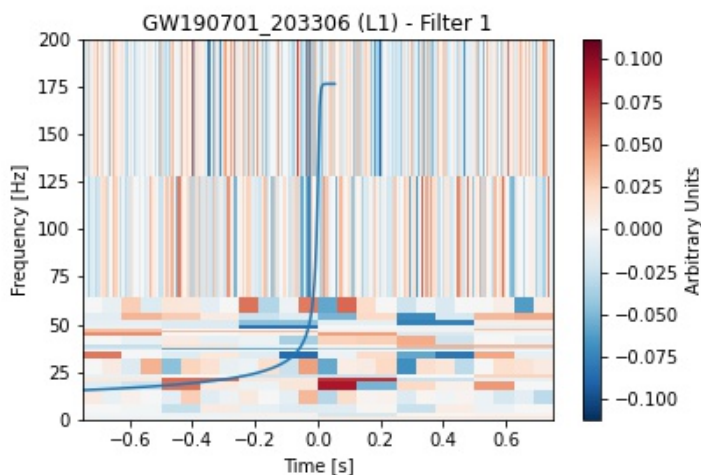
A.6 GW190701_203306

The last event evaluated by the developed code is also one of the O3a detections (ABBOTT et al., 2021). This gravitational wave event was generated by a binary black hole of masses $53.9 M_{sun}$ and $40.8 M_{sun}$ produced a black hole of mass $94.3 M_{sun}$ and released energy in GW form on 1st July 1019 at 20h33'06".

This particular case possess a lot of noise fluctuations around the time of the event, therefore it is expected to see tiles with an intense colour grouped in lower frequencies. Thus, the time interval used for this analysis was 8 and 16 seconds and only filtered data was used, so there would be no mixing of noise fluctuation and the representation of the noise that the data possess.

Figure A.9 displays the result of the GW190701_203306 (L1) event when using a time interval of 8 seconds, a Tukey Window shape parameter of 0.3 and the bandpass filter that attenuates frequencies out of the $[20,1000]$ Hz range. As it was expected, there are tiles grouped forming the noise configurations presented in LIGO's spectrogram.

Figure A.9 - The time-frequency representations for the GW190701_203306 (L1) event, when using $T = 8s$, with a bandpass filter ($[20,1000]$ Hz) and a sampling rate of 4kHz. In the plot the event's $f(t)$ is presented as the blue curve.



The result of the other attempt using a time interval of 16 seconds is presented in Figure A.10, where the maximum coefficient has module 0.1, and the grouped tiles are still present, which indicates that the code's results are solid with slight differ-

ences when the parameters changes. One last time the colorbar scale is presented on the right figure to perform the intensity comparison of the events.

Figure A.10 - The time-frequency representations for the GW190701_203306 (L1) event, when using $T = 16$ s, with a bandpass filter ([20,1000]Hz) and a sampling rate of 4kHz with different colorbar scales. In both plots the event's $f(t)$ is presented as the blue curve.

