
From Einstein to Einstein Telescope: From Testing Predictions to Addressing Future Challenges

Justin Janquart

ISBN:

DOI:

Printed by:

Copyright © 2023 Justin Janquart



This work originates as part of the research program of the Foundation for Fundamental Research on Matter (FOM), and falls as of April 1, 2017 under the responsibility of the Foundation for Nederlandse Wetenschappelijk Onderzoek Instituten (NWO-I), which is part of the Dutch Research Council (NWO).

From Einstein to Einstein Telescope: From Testing Predictions to Addressing Future Challenges

**Van Einstein tot Einstein Telescope:
van het testen van voorspellingen tot het aanpakken van toekomstige uitdagingen**

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op DAG XX MAAND 2023 TIJD
door

Justin Janquart
geboren op 12 juli 1997
te Geraardsbergen, België

First Promotor

Prof. dr. Chris Van Den Broeck

Second Promotor

dr. Sarah Caudill

Assessment Committee

Prof. dr. A

Prof. dr. B

Prof. dr. C

*Home is behind, the world ahead,
And there are many paths to tread
Through shadows to the edge of night,
Until the stars are all alight*

– J.R.R. Tolkien, The Fellowship of the Ring.

ACKNOWLEDGMENTS

Life is a journey, and each adventure has its chapter(s). My Ph.D. will always be a highlight of my story, providing many opportunities to learn and grow on intellectual, professional, and personal levels. Besides the particular experience, it also led me to meet many exceptional people.

First, I would like to thank Chris for making this journey possible. Your guided freedom has made it possible for me to realize so many things. Your kind advice has helped me navigate the research realm and learn so much on various topics. I truly hope to be able to pass on this experience with as much care and passion as you.

I would also like to thank Otto. You opened the door to the thrilling world of lensing. You allowed me to take on even more responsibilities and grow as a scientist in a worldwide collaboration and as a person. Besides, your general advice and help have been greatly appreciated, even when coming in the form of a Sunday morning Zoom meeting.

Tim and Anuradha, we have not seen each other much, probably due to COVID times, but you have certainly been part of this. The two of you have been working with me since day one when I started as a stressed newbie Ph.D. to work on overlapping signals. I have learned an enormous amount from the two of you. I look forward to continuing to collaborate with you.

The office has not been the most used place in the first half of my time here, but when being there, it has been a pleasure to share my office with you, Haris. I have always been amazed by your calm and patience when dealing with research. Since day one (before office time), you have been there to answer my questions, even the most trivial ones. Thanks a lot!

People often say that research is a tough world. I would not deny it, but one often forgets to mention the incredible solidarity existing between Ph.D. candidates. Therefore, I would particularly like to thank Anna, Harsh, Melissa, Stefano, and Tomek. More than simple collaborators, you have grown to be friends. Nothing of this would have been the same without you. From the laughs to the more serious discussions, from the simple meetings to the more adventurous travels, I all enjoyed it. Additionally, I would also like to thank the entire group – Booshan, Chinmay, Marc, Pawan, Peter, Quirijn, Sarah, and Soumen – for the various discussions and experiences we had together.

Tjonne, I would also like to thank you. Since you visited Utrecht in the summer, you have been a point of contact outside Utrecht University. Meeting and discussing with you in various conferences and contexts has been a great way to learn and see things from a different perspective.

I also thank the Gravitational-Wave group in Liège, particularly Jean-René, Grégory, Maxime, and Vincent. It is the group where everything started when you, Jean-René, accepted me to do my master's thesis with you. Then, you pushed me to explore new horizons and look outside of Belgium for my Ph.D. It turns out it was a life changer and has made me realize how much the World had to offer. Even after my master, we continued working together, and I continued to learn from all of you. *Pour tout cela,*

merci du fond du coeur.

Among all the incredible people I have had the pleasure to meet and discuss with, I would particularly like to thank Mick. Whilst we have not met much in person – probably due to the distance, you have been a great support with our mutual chat discussion and quick help channel. Although it pains me a bit to say so, you probably have a non-trivial contribution to the improvements I made in writing English, among other things. For that and everything else, *tapadh leat*.

I would also like to thank Alex. Our collaborations have been fruitful, and you have permitted me to learn much more about machine learning. *Dankjewel* for reaching out to us in the first place, but also for all the discussions about scientific and non-scientific topics. All of this would not have been the same without this. I hope we can pursue our crazy ideas in the future.

Additionally, I would like to thank Freija for the Dutch courses. I thought I would speak more Dutch coming to the Netherlands. The truth is that I spoke Dutch mostly with you in our various conversations. *Dankjewel* for getting some Dutch back into my head through nice conversations.

I could go on and on, listing all the people I have met and discussed with, who have made my journey nicer and more interesting. Instead, let me thank all my international friends and collaborators met along the way. I am sure those concerned will recognize themselves. You guys have also been a part of this, making it lighter and sometimes permitting me to release some of the stress.

I would also like to thank my family. You have always been there to chat with me, often telling me to take things a bit easier. I have not always realized how much this meant you cared about me. Thank you for sometimes bringing me back to Earth, and making me realize I should take a step or two back. We may not be a fairy-tale model family, but we are the best family I could have hoped for.

An extra member of my family, who joined after all of this started and made for important moral support: Newton, my cat. You will probably never read this (or you will, and I will not know), but you are a real stress reliever. Working from home has never been so smooth since you were here to come and sleep (on my desk or shoulder) close by, add some extra words in my codes and documents (authentic extract: *ofehmmmoivbeza*), or simply fool around when I was stuck in a once-more too-long Zoom call or on a bug.

Last but certainly not least, I would like to thank my wife, Amélie. One of the (maybe numerous) days I was complaining about something that was not working, you told me *"I am sorry, but I cannot help you there."*. However, you have been helping me since day one. Your patience, love, and care made it possible for me to achieve all of this. You raise me up to more than I can be. For that, and everything you do for me, *merci!*

External contributions

The results shown in this work have had different contributions and are often the result of collaborative efforts. Therefore, I would like to thank all my collaborators and

acknowledge their contributions. In particular, the work carried out by dr. Grégory Baltus for the results presented in Chapters 12 and 13, where we built the networks and performed the various tests out together. Another significant contributor is Alex Kolmus, who made the neural networks used in Chapter 14 and gave us access to an early version of his normalizing flow network and guided us in its adaptation to overlapping signals in Chapter 18.

CONTENTS

List of Figures	xvii
List of Tables	xxiii
List of Acronyms	xxv
1 A Brief History of Gravitational-Wave Physics	1
I Introduction to Gravitational Waves, Related Phenomenology, and Other Useful Tools	13
2 General Relativity and Gravitational Waves	15
2.1 The Main Ingredients of General Relativity	15
2.2 First Hints of Gravitational Waves	17
2.2.1 Linearized General Relativity	17
2.2.2 Let There be Waves	19
2.2.3 Basic Effects of Gravitational Waves on Matter	19
2.2.4 Premises of Detection	21
2.3 Generation of Gravitational Waves	22
2.3.1 Multipole Expansion	23
2.4 Gravitational-Wave Emission from Rotating Objects	25
2.5 Inspiring Objects	27
3 Waveforms: Key to Analysis of the Data	31
3.1 Modeling the Inspiral: the Post-Newtonian Formalism	32
3.2 Modelling the Full Signal: Numerical Relativity	34
3.3 Effective One-Body Waveforms	35
3.4 Phenomenological Waveforms	35
3.5 Higher-Order Modes in Gravitational Waves	40
4 Gravitational-Wave Data and its Analysis	43
4.1 The Noise and its Challenges	43
4.1.1 Extracting a Gravitational-Wave Signal from Noise	44
4.1.2 Wiener Filters for Gravitational-Wave Searches	46
4.2 Bayesian Statistics, the Key to Data Analysis	48
4.2.1 Bayes' Theorem	49
4.2.2 Summarizing the Posteriors	50
4.2.3 Hypothesis Testing	52
4.2.4 Final Note on Bayesian Statistics	53

4.3	Nested Sampling	53
4.3.1	Obtaining the Evidence and Posteriors with Nested Sampling	53
4.3.2	Nested Sampling for Gravitational Waves	57
4.4	Markov Chain Monte Carlo	57
5	Deflection of Waves by Massive Objects	61
5.1	Gravitational-Wave Lensing	64
5.1.1	Geometric Optics Limit	66
5.1.2	Wave Optics	67
5.2	Lens Models	68
5.2.1	Point Mass Lens	68
5.2.2	Singular Isothermal Sphere	69
5.2.3	Singular Isothermal Ellipsoid	69
5.3	Searching for Strongly-Lensed Events in the Data	70
6	Machine Learning Concepts	77
6.1	The Building Blocks	77
6.1.1	The fundamental Entity: the Neural Network	77
6.1.2	Loss Functions, the Key to Training	79
6.1.3	Gradient Descent, the Key to a Working Network	80
6.1.4	Better Explore the Loss Function Space: Optimizers	82
6.1.5	Overview of a Basic Training Process	83
6.2	Finding Spatial and Time Correlation in Data	84
6.2.1	Convolutional Neural Networks	84
6.2.2	Making Deeper Networks: Residual Connections	85
6.2.3	Batch Normalization	87
6.3	Neural Networks for Probability Density Estimation	87
6.3.1	Transforming Simple to Complex Probability Distributions	87
6.3.2	Continuous Normalizing Flows	90
6.3.3	Accounting for Data in Normalizing Flows	92
6.4	Pre- and Post-Processing the Data	92
6.4.1	Normalizing the Data	93
6.4.2	Extracting Essential Features of the Data	93
6.4.3	Linking Machine Learning With Classical Methods	95
II	Gravitational-Wave Lensing: Magnifying General Relativity	97
7	Searches for Strong Gravitational-Wave Lensing and Related Issues	99
8	GOLUM: A Fast and Precise Methodology to Search for and Analyze Strongly Lensed Gravitational Waves	101
8.1	Recasting the Lensed Evidence for Fast Computation	101

8.2	Evaluating the Conditioned Evidence	103
8.3	Additional Likelihood Evaluation Speed up	104
8.4	Example Analysis	105
8.5	Adaptation to Multiple-Image Analyses	110
8.6	Quadruple Image Analysis: Sky Localization	111
8.7	A Note on Selection Effects	113
8.8	Concluding Remarks Concerning GOLUM	113
9	On the Identification of Individual Gravitational-Wave Image Types of a Lensed System Using Higher-order Modes	115
9.1	Methodology	115
9.2	Results	116
9.2.1	Type I–Type II Systems	117
9.2.2	Type I–Type I and Type II–Type II Systems	117
9.2.3	Using Templates Without HOMs to Analyze Systems with HOMs	118
9.2.4	Improved Probing of HOMs with Lensing	119
9.3	Summary and Conclusions	119
10	Ordering the Confusion: a Study of the Impact of Lens Models on Gravitational-Wave Lensing Detection Capabilities	121
10.1	Including the Lens Statistics in the Coherence Ratio	121
10.2	Injections and Setup of the Study	123
10.2.1	Binary Black Hole Population	123
10.2.2	Population Analyses	124
10.2.3	Determining Lensed Candidates	127
10.3	Results	128
10.3.1	From Posterior Overlap to Joint Parameter Estimation	129
10.3.2	Including the Correct Model	131
10.3.3	Using Other Models	132
10.3.4	Using Only the Time Delay	133
10.3.5	Analyses Using the Toy Models	137
10.4	Discussion on Triply and Quadruply Lensed Images	140
10.5	Conclusions and Outlook	141
III	Machine Learning: A Tool for Gravitational-Wave Data Analysis	145
11	Multi-Messenger Astronomy and the Need for Faster Analyses	147

12 Convolutional Neural Networks for the Detection of the Early Inspiral of a Gravitational-Wave Signal	149
12.1 Setup of the Study	150
12.1.1 Quantifying the Signal Present	150
12.1.2 Defining Different BNS Categories	152
12.1.3 Data Generation	154
12.2 Methodology	155
12.2.1 Convolutional Neural Network Architecture	155
12.2.2 Training and Testing our Neural Networks	156
12.3 Results and Discussion	158
12.3.1 Performance of the CNNs	159
12.3.2 Test on a Realistic BNS Population	161
12.3.3 Basic Curriculum Learning Exploration	162
12.4 Conclusions	163
13 Convolutional Neural Network for Gravitational-Wave Early Alert: Going Down in Frequency	165
13.1 Method and Setup	166
13.1.1 Data and Training Strategies	166
13.1.2 Description of the Neural Network	170
13.2 Results and Discussion	171
13.2.1 Performance of the Network	171
13.2.2 Estimation of the Number of BNS Inspirals Detectable in O4	174
13.3 Conclusions	177
14 Swift Sky Localization of Gravitational Waves using Deep Learning Seeded Importance Sampling	179
14.1 Methodology	180
14.1.1 Going from Simple Distributions to Complex Posteriors	181
14.1.2 Neural Model	182
14.2 Setup for the Experiments	184
14.2.1 Setup for Training and Evaluating the Neural Network	184
14.2.2 Setup for Importance Sampling	186
14.2.3 Generating Skymaps	187
14.3 Results	187
14.3.1 Importance Sampling	188
14.3.2 Generating Skymaps	188
14.4 Conclusions	190

IV Third-Generation Detectors: Interesting Science, Tough Analyses	195
15 Third-Generation Detectors: Expanding our Horizons	197
16 Biases in Parameter Estimation From Overlapping Gravitational-Wave Signals in the Third Generation Era	199
16.1 Overlap Rate Estimates	200
16.1.1 Methodology	200
16.1.2 Overlap Estimates	203
16.2 Parameter Estimation Setup	206
16.3 Results	211
16.3.1 Overlap of a BNS Signal with a BBH Signal	211
16.3.2 Overlap of Two BBH Signals	212
16.3.3 Overlap of Two BNS Signals	213
16.4 Impact of the Noise Realizations	215
16.5 Conclusions and Outlook	220
17 Parameter Estimation Methods for Analyzing Overlapping Gravitational-Wave Signals in the Third-Generation Detector Era	225
17.1 Description of the Methods	227
17.1.1 Joint Parameter Estimation	227
17.1.2 Hierarchical Subtraction	229
17.2 Setup of the Analyses	230
17.3 Results and Discussions	233
17.3.1 Hierarchical Subtraction	233
17.3.2 Joint Parameter Estimation	238
17.4 Zero-noise Results	242
17.4.1 Hierarchical Subtraction	242
17.4.2 Joint Parameter Estimation	244
17.5 Conclusions	247
18 Normalizing Flows as an Avenue to Study Overlapping Gravitational-Wave Signals	251
18.1 Machine Learning for Overlapping Gravitational Waves	251
18.2 Data and Setup	254
18.3 Results	255
18.4 Conclusions and Perspectives	259
V Concluding Remarks and Future Perspectives	261
Public Summary	269

Openbare Samenvatting	277
Curriculum Vitae	286
Bibliography	289

LIST OF FIGURES

1.1	Pictures of the LIGO and Virgo detectors and an artist representation of Einstein Telescope	4
1.2	Simplified representation of a GW interferometer.	5
1.3	Time-frequency maps and sky location for GW170817.	6
1.4	Representation of all the CBCs observed by the LVK collaboration after the third observing run.	8
1.5	Horizon distance for the planned 3G detectors.	10
1.6	Sensitivity curves for the current and future GW detectors.	10
2.1	Representation of the GW polarization evolution over a period.	20
2.2	Representation of the angles linking the GW-frame and detector-frame axes.	22
2.3	Representation of a binary system made of two point-mass particles on a circular orbit.	25
3.1	Illustration of the spins' impact on a GW waveform.	34
3.2	Representation of the absolute value of different spin-weighted spherical harmonics.	40
4.1	Representation of the mean O3 noise PSDs for the two LIGO and Virgo detectors and the glitch present in the GW170817 LIGO-Livingston data.	45
4.2	Representation of the template bank used for the GstLAL search in the second observing run.	47
4.3	Representation of joint and marginalized posterior distributions with their confidence intervals for the sky location of a GW event.	51
4.4	Illustration of the nested sampling process.	54
4.5	Illustration of the Metropolis-Hastings algorithm.	58
5.1	Representation of the microlensing and strong-lensing effects on a GW signal.	62
5.2	Illustration of the virtual detector network obtained thanks to strong lensing.	63
5.3	Representation of the situation considered to derive the amplification factor for GW lensing.	65
5.4	Illustration of the distortion due to type II image for strongly-lensed GW signals with an important higher-order mode contribution.	73
5.5	Strong lensing false-alarm probability with and without a lens model as a function of the observation time.	75
6.1	Illustration of a simple neural network, a perceptron, and a neuron with the link between them.	79
6.2	Representation of a convolution layer in a neural network.	85
6.3	Representation of a residual connection.	86

8.1	Posterior distribution for the relative lensing parameters for an example lensed pair analyzed using GOLUM	107
8.2	Posterior distribution for the sky location for an examples lensed pair analyzed using GOLUM	108
8.3	Posterior distribution for the relative magnification and time delay for a pair of lensed images with a sub-threshold secondary image.	109
8.4	90% credible region of the sky location for a lensed singlet, pair, triple, and quadruplet.	112
9.1	Posterior probability values for the recovery of the Morse factor for a lensed pair (made of a type I and a type II image) and a single type II image as a function of the total higher-order mode content.	118
9.2	Posterior distributions for the inclination and mass ratio for an unlensed event and a lensed type I-type II image pair.	120
10.1	Time delay and relative magnification distribution for the \mathcal{H}_t , $\mathcal{M}_{\mu,t}$ and $\mathcal{W}_{\mu,t}$ models.	126
10.2	Performance comparison between posterior overlap and GOLUM	130
10.3	Comparison of the background and foreground coherence ratio with and without the inclusion of the true lens model.	132
10.4	Comparison of the foreground and background values for the $\mathcal{C}_{\mathcal{M}_{\mu,t}}$ and $\mathcal{C}_{\mathcal{W}_{\mu,t}}$ detection statistics.	134
10.5	ROC curves for the detection statistics including the correct and wrong density profiles.	135
10.6	Comparison of the detection statistic for the background and foreground when no model and the correct model with and without relative magnification are included.	136
10.7	Comparison between the foreground and background detection statistics for a wrong density profile and when no model is included in the detection statistics.	136
10.8	Comparison of the detection statistics for the unlensed and lensed events for the toy models A and B, and the coherence ratio.	138
10.9	Comparison of the detection performance for the correct lens density profile for different bounds taken on the distributions of the lensing parameters.	139
10.10	Comparison of the detection performance for the lens models coming from existing lensing statistics catalogs with and without including relative magnification.	143
12.1	Evolution of the partial-inspiral SNR as a function of the fractional duration of the inspiral.	151
12.2	Full and partial BNS template with its frequency evolution.	152

12.3	Representation of the noise and the GW signal injected in it before whitening.	155
12.4	Architecture of the best-performing neural network for our early-alert system.	156
12.5	SNR and partial-inspiral SNR distributions for the training and testing sets.	158
12.6	True alarm probability as a function of the distance and partial-inspiral SNR for our three neural networks.	159
12.7	True alarm probability as a function of the distance for different observational time windows.	160
12.8	Missed and detected BNS inspiral signals with their corresponding partial-inspiral SNR in five years of data for a LIGO detector at design sensitivity.	162
12.9	True alarm probability as a function of the distance with and without curriculum learning.	163
13.1	Representation of the different PSDs considered in Chapter 13.	167
13.2	Example noise realizations considered in Chapter 13	167
13.3	Representation of the noise artifacts present in the real O3 data and the corresponding vetoed data.	169
13.4	True alarm probability as a function of the mean partial-inspiral SNR for different maximum inspiral frequencies and different noise realizations.	172
13.5	Time left before the merger as a function of the distance for the early detection of a GW170817-like signal.	174
13.6	Representation of a BNS signal injected in noise with the corresponding triggers given by our network as we slide a window through the data.	175
13.7	Chirp mass and time before the merger as a function of the network SNR for the BNS signals detected early by our neural network in five years of simulated O4 data.	177
14.1	Graphical depiction of the neural network developed for swift sky location.	183
14.2	SNR sampling distribution for the training and validation data.	186
14.3	Characterization of the neural network's sky location accuracy on the test set depending on the SNR.	189
14.4	Mean relative error on the component masses estimated by the neural network as a function of the optimal SNR.	190
14.5	Evolution of the precision of importance sampling with the number of samples generated for different signal loudness.	191
14.6	Log-likelihood values for the true right ascension for BILBY and importance sampling as a function of the number of samples considered and for different SNRs.	192
14.7	Comparison of skymaps predicted by our neural network with and without applying importance sampling and predicted by BILBY.	193

16.1	BBH merger rate density according to Oguri's fit when including population I, II, and III stars.	201
16.2	Fraction of detected BNS signals with a given number of CBC mergers taking place while the signal is in band.	205
16.3	Individual waveforms and overlap scenarios considered in the study of biases for overlapping signals.	208
16.4	Posterior distributions for BNS and BBH intrinsic parameters when they are overlapped with each other with various differences in merger times. .	210
16.5	Posterior distributions for the total mass and mass ratio when high and low mass BBH signals are overlapped.	213
16.6	Posterior distributions for intrinsic BNS parameters when another BNS signal merges at the same time or two seconds earlier.	214
16.7	Posterior distributions for intrinsic BNS parameters when overlapped with BBH signals for different noise realizations.	216
16.8	Posterior distributions for BBH parameters when overlapped with a BNS signal for different noise realizations.	217
16.9	Posterior distributions for BBH parameters when overlapped with another BBH signal for different noise realizations.	218
16.10	Posterior distributions when two BNS signals overlap for different noise realizations.	219
17.1	Representation of two overlapping BBH signals merging 0.1 s apart with two different SNR ratios.	232
17.2	Comparison of the recovered posteriors for the hierarchical subtraction and single parameter estimation methods, with and without noise.	234
17.3	Mismatch between the injected and recovered waveforms for the single parameter estimation analysis versus the same quantity for hierarchical subtraction when the injections are done in noise.	235
17.4	Comparison of the offset of the recovered posteriors for single parameter estimation and hierarchical subtraction with noise.	236
17.5	Comparison of the width of the recovered posteriors for single parameter estimation and hierarchical subtraction with noise.	236
17.6	Comparison of the mismatch between injected and recovered waveforms and the offset of the posteriors for the first and third hierarchical subtraction steps.	237
17.7	Comparison of the recovered posteriors for the joint parameter estimation and single parameter estimation methods, with and without noise.	239
17.8	Comparison of the mismatch between the injected and recovered waveforms for joint parameter estimation versus single parameter estimation, and joint parameter estimation versus hierarchical subtraction, with noise. .	240
17.9	Comparison of the offset of the recovered chirp mass for joint parameter estimation versus single parameter estimation, and joint parameter estimation versus hierarchical subtraction, with noise.	241

17.10 Comparison of the width of the recovered chirp mass posteriors for single parameter estimation and joint parameter estimation, with noise.	242
17.11 Mismatch between the injected and recovered waveforms for the single parameter estimation method versus the hierarchical subtraction analysis, without noise.	243
17.12 Offsets of the recovered chirp mass posteriors for single parameter estimation versus hierarchical subtraction, without noise.	243
17.13 Widths of the recovered chirp mass posteriors for single parameter estimation versus hierarchical subtraction, without noise.	244
17.14 Mismatch between the injected and recovered waveforms for joint parameter estimation versus single parameter estimation, and joint parameter estimation versus hierarchical subtraction, without noise.	245
17.15 Comparison of the offset of the recovered chirp mass for joint parameter estimation versus single parameter estimation, and joint parameter estimation versus hierarchical subtraction, without noise.	246
17.16 Normalized width of the 90% confidence interval for the joint parameter estimation method versus the single parameter analysis, without noise. . .	247
18.1 Graphical depiction of our machine-learning analysis framework for overlapping BBH signals.	254
18.2 Example recoveries using our machine-learning approach for the sky location and component masses of two overlapping BBH signals.	256
18.3 P-P plots for the parameters recovered by our machine-learning approach for two overlapping BBH signals.	257
18.4 Comparison between the posteriors recovered using our machine-learning approach and using the Bayesian joint parameter estimation analysis. . .	258

LIST OF TABLES

1.1	Overview of the outstanding events seen during the third LVK observing run.	7
8.1	Summary of the injection parameters used for the example lensed image-pair analyses.	106
10.1	Summary of the different lens models used to study the impact of the inclusion of a lens model on the strong-lensing false-alarm probability. . .	127
10.2	Summary of the different models used in Chapter 10 with their associated false-alarm probability.	142
12.1	Summary of the three BNS categories used in our early-alert system with their corresponding characteristics.	154
12.2	Detailed architecture of the best-performing neural network for our early-alert system.	157
13.1	Overview of the different data sets used when performing curriculum learning over the signals' maximum frequency.	170
13.2	Summary of the characteristics of the events detected early in five years of simulated O4 data.	176
14.1	Priors used for the data generation when training and testing our neural network.	185
16.1	Number of mergers detected by a network of two CE and one ET in one year of simulated data for different SNR thresholds.	204
16.2	The number of seconds in a year with at least two mergers occurring, depending on their types.	204
16.3	Typical number of CBC mergers happening while a BNS signal is in band. .	205
16.4	Number of CBC signal pairs with both SNRs between 15 and 30, and such that their mergers occur within 2 seconds or less of each other. . .	206
16.5	Values of the KS statistic comparing the recovered probability density distributions for BNS signals with and without overlapping BBH signals for different merger time differences.	209
16.6	Value of the KS statistic comparing the recovered posteriors for BBH parameters with and without an overlapping BNS signal.	211
16.7	Values of the KS statistic comparing the recovered posterior distributions for BBH parameters with and without another BBH signal overlapping. .	214
16.8	Values of the KS statistic comparing the recovered posterior distributions for BNSs with and without another BNS signal overlapping.	215

16.9	Values of the KS statistic comparing the recovered posterior distributions for BNS parameters with and without an overlapping BBH signal for different noise realizations.	220
16.10	Values of the KS statistic comparing the posterior distributions for BBH parameters with and without an overlapping BNS signal for different noise realizations.	221
16.11	Values of the KS statistic comparing the recovered posterior distributions for a BBH signal with and without another BBH signal overlapping for different noise realizations.	221
16.12	Values of the KS statistic comparing the recovered posterior distributions for BNS parameters with and without another BNS signal overlapping for different noise realizations.	222
17.1	Overview of the probability distributions used to generate the BBH population and those used as priors for the parameter estimation analyses.	230
18.1	Summary of the functions used to generate the BBH parameters for the overlapping signals analyzed with our machine-learning framework.	255

LIST OF ACRONYMS

- **ALF**: asynchronous leapfrog
- **BBH**: binary black hole
- **BBO**: Big Bang Observatory
- **BH**: black hole
- **BNS**: binary neutron star
- **CBC**: compact-binary coalescence
- **CCDF**: complementary cumulative distribution function
- **CDF**: cumulative distribution function
- **CE**: Cosmic Explorer
- **CNF**: continuous normalizing flow
- **CNN**: convolutional neural network
- **DL**: deep learning
- **EM**: electromagnetic
- **EoS**: equation of state
- **ET**: Einstein Telescope
- **FAP**: false-alarm probability
- **FAR**: false-alarm rate
- **FFT**: fast Fourier transform
- **FN**: false negative
- **FP**: false positive
- **FPP**: false-positive probability
- **GELU**: Gaussian error linear unit
- **GR**: general relativity
- **GRB**: gamma-ray burst
- **GW**: gravitational wave

- **HOM**: higher-order mode
- **HS**: hierarchical subtraction
- **IMBH**: intermediate-mass black hole
- **ISCO**: inner-most stable circular orbit
- **JPE**: joint parameter estimation
- **KDE**: kernel density estimator
- **KL** (divergence): Kullback-Leibler (divergence)
- **KS** (statistic): Kolmogorov-Smirnov (statistic)
- **LIGO**: laser interferometer gravitational-wave observatory
- **LISA**: laser interferometer space antenna
- **LVK**: LIGO-Virgo-KAGRA
- **maae**: mean absolute angular error
- **MALI**: memory effective asynchronous leapfrog
- **MCMC**: Markov chain Monte Carlo
- **ML**: machine learning
- **MMA**: multi-messenger astronomy
- **MSE**: mean-square error
- **MVG**: multi-variate Gaussian
- **NF**: normalizing flow
- **NLL**: negative log-likelihood
- **NN**: neural network
- **NR**: numerical relativity
- **NS**: neutron star
- **NSBH**: neutron star-black hole
- **ODE**: ordinary differential equation
- **OTW**: observational time window

- **PDF**: probability density function
- **PI SNR**: partial-inspiral signal-to-noise ratio
- **PN**: post-Newtonian
- **PSD**: power-spectral density
- **ReLU**: rectified linear unit
- **ROC**: receiver-operator characteristics
- **SIE**: singular isothermal ellipsoid
- **SIS**: singular isothermal sphere
- **SNR**: signal-to-noise ratio
- **SPE**: single parameter estimation
- **SPEMD**: singular power-law elliptical mass distribution
- **SVD**: single-value decomposition
- **TAP**: true-alarm probability
- **TN**: true negative
- **TP**: true positive
- **TT** (gauge): traceless-transverse (gauge)
- **VMF** (distribution): Von Mises-Fisher (distribution)

1

A BRIEF HISTORY OF GRAVITATIONAL-WAVE PHYSICS

For years, Newtonian gravity was used to describe the planets' motions. However, it had some shortcomings, such as the impossibility to explain the advance of Mercury's perihelion. Therefore, and based on the observation that Newton's gravitational laws are incompatible with Maxwell's equations for electrodynamics, Einstein proposes his *Special Relativity* [1] in 1905. Later, based on this first step to a novel theory for gravity, Einstein postulates his theory of *General Relativity* (GR) [2]. Einstein's depiction of gravity is fundamentally different from Newton's since the gravitational force is replaced by the curvature of the spacetime manifold. However, despite the different bases of the theories, in the case of weak curvature and low velocities, GR reduces to Newtonian mechanics.

Since its description by Einstein, GR has faced many tests, passing each of them. The first test for GR was to explain the precession of Mercury's perihelion. Deriving the planet's motions in GR, one finds that its orbital motion is an ellipse with a small deviation corresponding to the precession effect. More than just solving existing problems, GR also makes some predictions of new phenomena one can expect. Among other examples, GR predicts the gravitational redshift of light, the deflection of light around massive objects — called *gravitational lensing* —, the existence of black holes (BHs), and the emission of gravitational waves (GWs). The observation of these phenomena would add to the validity of GR. Some of these predictions were already made by Einstein (like lensing and GWs). However, he was often confused by them, mainly because the objects known at that time were not dense enough to lead to observable phenomena.

Giving more details on the above examples in turn, when an electromagnetic (EM) wave or a photon travels out of (resp. in) a gravitational potential it seems to lose (resp. gain) energy, corresponding to a decrease (resp. increase) in the wave frequency, leading to gravitational redshift (resp. blueshift). While this effect was already predicted by Einstein when he published his full theory of relativity, it was observed for the first time in the *Pound-Rebka experiment* in 1959, by looking at the change in energy from a gamma-ray photon between its observation at the top and the bottom of a tower [3].

A bit earlier, in 1919, Eddington used a solar eclipse to measure the deflection of light due to massive objects. His observation was that the deflection angle was twice that predicted by Newtonian gravity, corresponding to the value predicted by GR [4]. Nowadays, the lensing of EM signals is commonly used in astronomy to study our Universe. For example, it is used to discover exoplanets [5, 6], probe dark matter [7–

[9], study the cosmological parameters [10–13], and investigate the nature of distant objects [14, 15]. As will be discussed later in this thesis, lensing is not unique to EM signals, GWs may also experience it.

In the 1960s, one found the first evidence of the existence of BHs – crucial for testing GR – thanks to the development of radio astronomy. These astrophysical objects are so dense that within a certain distance of their center, called the *event horizon*, nothing, not even light, can escape (hence the name). In a BH, one can find a location with infinite density, called the *singularity*. Since infinite density does not make much sense, this prediction from GR is often seen as the breaking point of the theory, and extra motivation to develop quantum-gravity theories. Such objects can be formed in different ways, the details of which have not yet been fully discovered. Depending on their mass, they could originate from the collapse of a star, the accretion of additional matter (including other BHs) by a BH, or even overdensities in the primordial Universe. While the first hints to the presence of BHs were determined by the impact they had on stars' motion in their vicinity, we are now in an exciting era where the first images of a BH have been made¹ [16–20].

Before detailing the next GR prediction mentioned previously, namely GWs, let us first recall that other dense objects exist in the Universe. Important objects for the development of GR are neutron stars (NSs). They originate from star collapse and are objects with such high density that the matter they contain is a kind of plasma where protons and neutrons are squeezed together very tightly, making an important laboratory to study the interactions under extreme conditions. In addition, these objects can have important magnetic and electric fields on top of the large gravitational field induced by their density. As we will see later, these objects play a role in the emission of GWs.

Finally, as a consequence of GR, when a massive object is accelerated, the resulting distortions of spacetime create a wave propagating at the speed of light in the fabric of spacetime. These are GWs. This was already known by Einstein [21]. However, the lack of knowledge about compact objects led him to believe this would never be detected. In principle, sources of GW emission are numerous. One can think of stars exploding asymmetrically, an NS with a mountain rotating on itself, binaries made of compact objects, ... The latter is the most promising as the GW emission is a faint effect, and in particular, even for very dense and massive objects such as orbiting BHs and NSs, the GW amplitude remains extremely small. However, this does not mean that the detection is impossible.

The first step to prove the emission of GWs was made by indirect measurement. The observations of the *Hulse-Taylor binary* in 1975 [22], consisting of a pulsar² and an NS enabled Hulse and Taylor to show a decay in the orbital period. Confronting this decay with the prediction of GR accounting for GW emission showed that the two were in

¹The image was obtained through the Event Horizon Telescope, and it corresponds to a massive BH at the center of the M87 galaxy [16–20].

²Pulsars are rapidly spinning NSs emitting a jet which can periodically point towards Earth and be detected.

good agreement. Therefore GW emission would be expected to be occurring within this system³ [23].

In addition to indirect detections, one would also like to make a direct detection of GWs. To accomplish this, the development of precision instruments started as early as the 1960s, with two main types of detectors being considered: *resonant-mass detectors* and *interferometers*. The first is a solid object resonating with the passage of a GW. So far, they have not led to any detection [24]. The other type of detectors has been more successful as they correspond to the design followed by the LIGO (Laser Interferometer Gravitational-wave Observatory) [25] and Virgo [26] detectors, which were at the origin of the observations kick-starting GW astronomy. Areal pictures of the LIGO-Hanford⁴ and Virgo detectors are shown in the top row of Fig. 1.1. While the basic design of these detectors was quite well settled in the 1970s, a lot of work still needed to be done before detection was possible. In principle, the idea is to send a laser onto a mirror, which splits it in between the two arms. It is then reflected in the readout. If the two arms have the same length, no light is observed as the two wavefronts interfere destructively; else, light is observed. The passage of a GW would squeeze and stretch the arms differentially, leaving an imprint in the recorded data. Fig. 1.2 represents a simplified design for such a detector.

To even think about detecting a GW signal from compact binary coalescences (CBCs, binaries made of two compact objects like BHs and NSs), one has to have the capacity to measure a relative difference in arm-length $\frac{\Delta L}{L} \sim 10^{-21}$ [24]. To achieve such sensitivity, important efforts were made in identifying the noise sources. Some of the implementations enhancing the simple Michelson interferometer are the addition of Fabry-Pérot cavities — which extend artificially the length of the detectors, the use of power-boosted lasers, and the operation at high vacuum to avoid interference with dust particles [24]. One also wants to dodge external noise such as seismic activities. If one can detect motions small enough to find GWs, one can also detect the motion of the tectonic plates or human activities. This is avoided by suspending the mirrors to a system of pendulums and using active damping [24]. An additional way to isolate the detector is to go underground, as has been done for the Japanese KAGRA detector [27–29], and is planned for the upcoming Einstein Telescope (ET) [30, 31]. The bottom row of Fig. 1.1 shows an artist representation of ET.

To finally get to the first direct GW observation, years of development have been necessary, both on the theoretical and the technological side. Owing to these upgrades, in September 2015, the first GW event ever was detected: GW150914⁵ [33], seen in coincidence by the two LIGO detectors (LIGO-Hanford and LIGO-Livingston). This

³This discovery led to a Nobel Prize for Hulse and Taylor in 1993.

⁴There are two LIGO detectors. One is situated in Hanford and the other in Livingston. They are built with the same design and technology.

⁵GW events detected by the LIGO and Virgo collaboration follow the naming convention GWYYMM-DD, where YY is the year, MM the month, and DD the day, up to the third observing run. From there on, it is also followed by the UTC observation time.

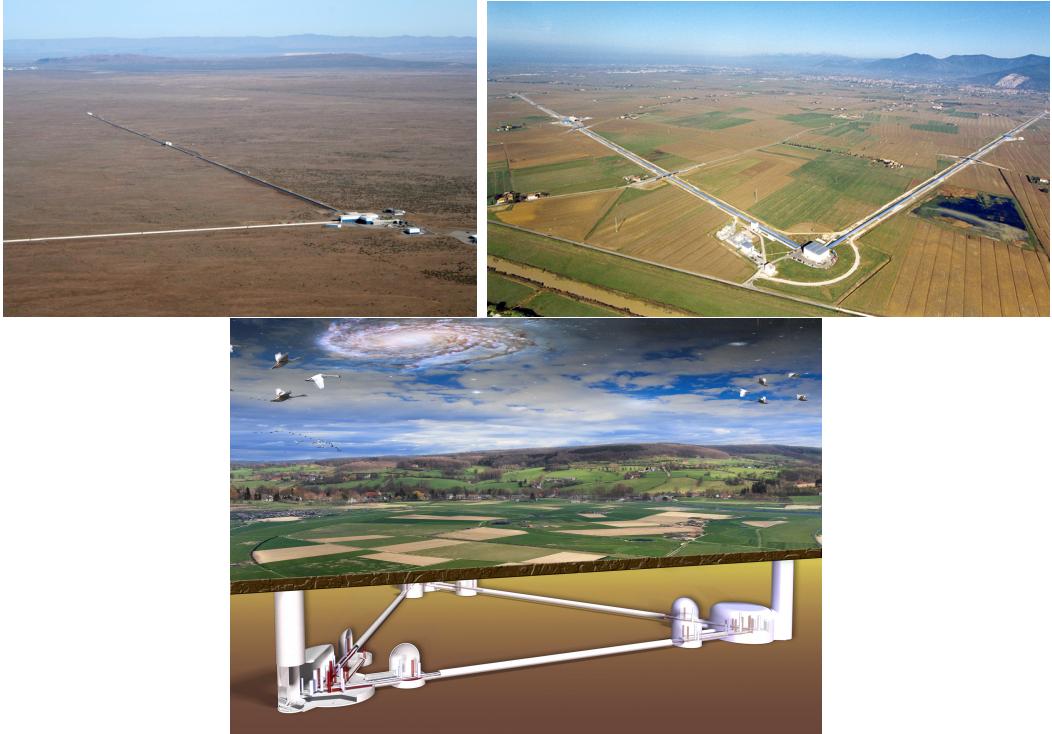


Figure 1.1: *Upper left:* Areal picture of the LIGO-Hanford detector (credits: Caltech/MIT/LIGO Lab). *Upper right:* Areal picture of the Virgo detector (credits: Virgo Collaboration). *Bottom:* Artist representation of Einstein Telescope (credits: Nikhef).

observation of two merging BHs represents a breakthrough for GR science⁶. It is the first direct observation of GWs, but also additional proof of the existence of BHs and binary BHs (BBHs), showing that these objects can form and make binaries in a time smaller than the age of the Universe. The nature of the objects present in this binary also enabled the scientific community to study the behavior of GR in the strongest regime ever detected, not finding any deviation [34].

While this first observation already represented a major discovery, it still needed to be cross-validated. By August 2017, 7 BBH mergers had been observed by the two LIGO detectors [35], and the Virgo detector joined the detector network. Soon after, GW170814 is detected as the first triple detection [36]. Since GWs are localized by comparing the time and phase upon arrival between the interferometers, the addition of a third detector leads to a major improvement in the event's sky location. In addition, it enables one to probe more than two polarizations and perform novel tests of gravity [36]. Regardless of the novelties brought by this triple detection, it is the GW170817 event [37] that makes history. It is identified as a binary NS (BNS) merger and was observed for about 100 s by the LIGO detectors. A representation of the signal's time-frequency map

⁶In 2017, Rainer Weiss, Barry Barish, and Kip Thorne were awarded the Nobel Prize for their contribution to the direct detection of GWs.

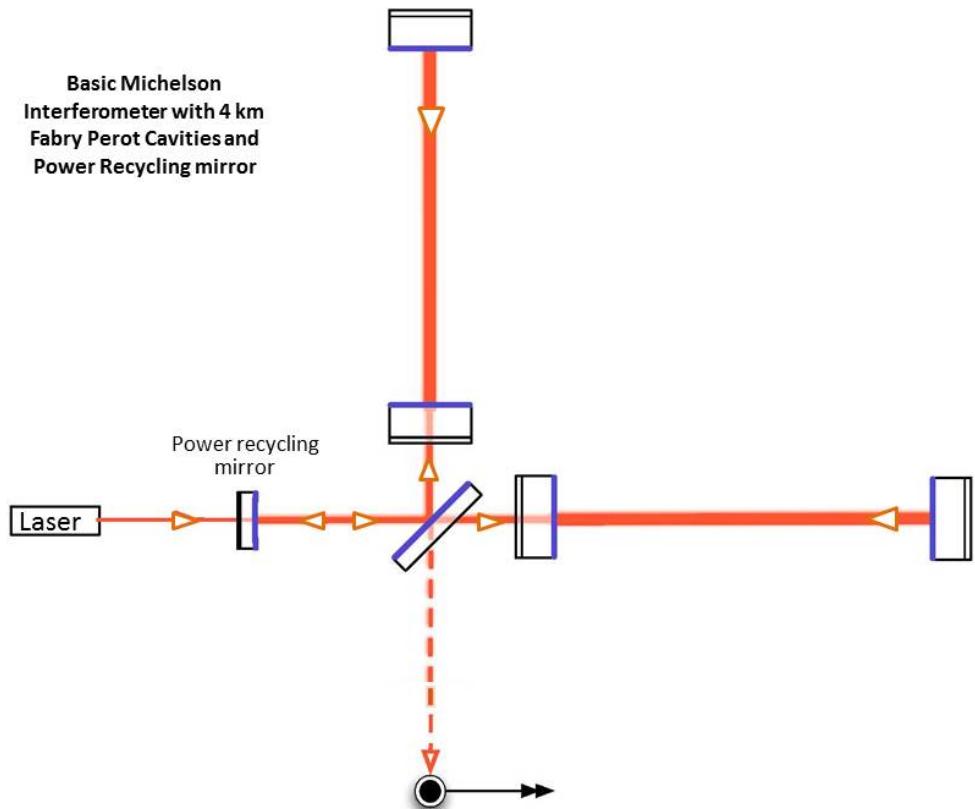


Figure 1.2: Simplified representation of a GW interferometer. The laser is sent to a beam-splitting mirror and gets reflected before being reconstructed. If the arms have the same length, the reconstruction is destructive, else it is constructive. The arm-lengths are artificially increased by the use of recycling mirrors so that the light bounces back several times in the cavity. Other enhancements have been implemented to reduce the noise sources and enable the first detection of GWs. Illustration taken from [32].

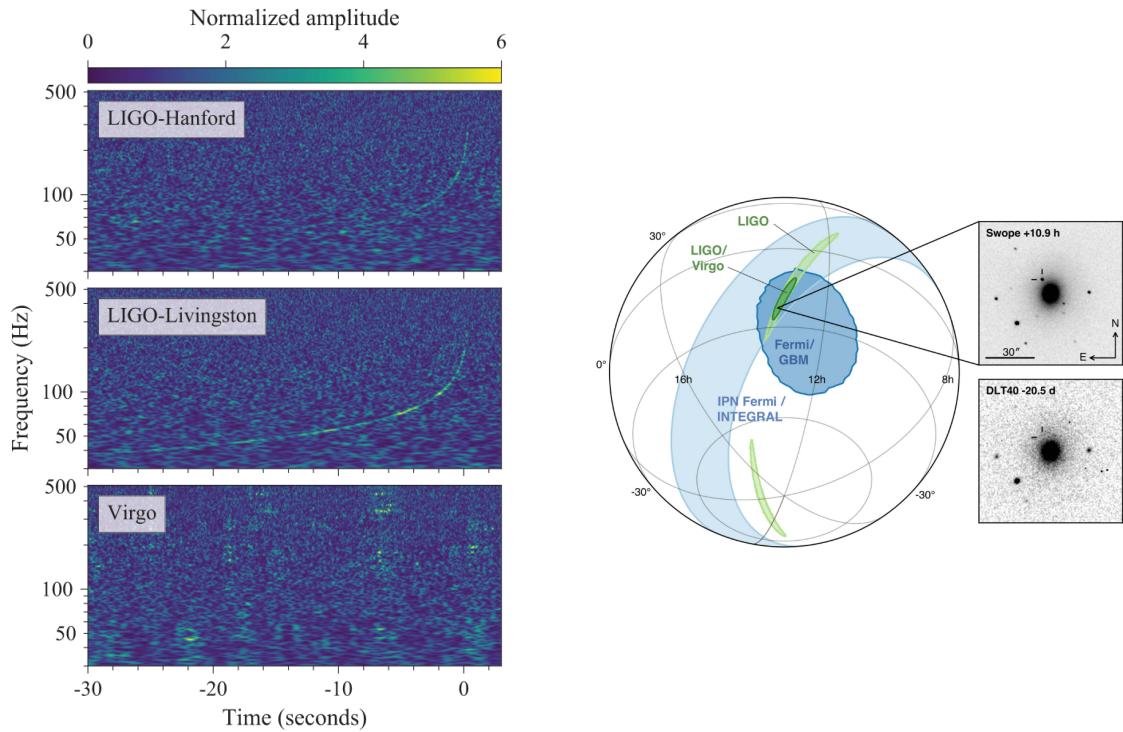


Figure 1.3: *Left:* Representation of the time-frequency map for the GW170817 event. One sees a clear signature in the two LIGO detectors. The signal is absent from the Virgo detector, reducing the possible sky location to blind spots of the detector [37]. *Right:* 90% confidence sky location for the GW, gamma ray, and optical signals and a picture of the host galaxy (NGC 4993) before and after the merger [38].

in the different detectors is given in the left panel of Fig. 1.3. Despite its loudness, it was not seen by the Virgo detector, which was online at that time. This absence of detection helped reduce the sky location even more, by reducing the possible locations to Virgo's blind spots (see the right panel in Fig. 1.3). The masses could be inferred from the GW data and showed agreement with most of the predicted NS equation of states (EoSs) [37].

Contrary to BBH mergers, BNS mergers are expected to emit EM radiation. Therefore, after the public alert for GW170817 was issued, astronomers made a global effort to find a counterpart to the merger. Thanks to the reduced sky location offered by the GW detector network, and by cross-correlating with possible EM sources, it was possible to link the BNS merger to a short gamma-ray burst (GRB). Afterward, a counterpart was found in all other EM bands. In the ultraviolet, optical and infrared bands, the observation of the kilonova⁷ led to the identification of the host galaxy: NGC 4993, while the observation of the radio and X-ray emission helped study the binary's environment [38]. The joint detection of GW and EM information provides new tests

⁷A transient surge of EM radiation about 1000 times brighter than a nova.

Event name	Brief description
GW190412	Asymmetric masses ($\sim 30M_{\odot}$ and $\sim 8M_{\odot}$) leading to the observation of higher harmonics [41].
GW190425	Low component masses (ranging from 1.12 to $2.52 M_{\odot}$), consistent with NS masses, but heavier than any other NS observed [42].
GW190521	Unusually high masses ($\sim 85 M_{\odot}$ and $\sim 66 M_{\odot}$), one of the components is probably in the mass gap, and the final BH is an intermediate-mass BH [43]
GW190814	Asymmetric masses ($\sim 23 M_{\odot}$ and $\sim 2.6 M_{\odot}$) with the observation of higher harmonics, and no confident identification of the nature of the lightest component [44].
GW200105	Observation of a NS-BH merger ($\sim 8.9 M_{\odot}$ and $\sim 1.9 M_{\odot}$) [45].
GW200115	Observation of a NS-BH merger ($\sim 5.7 M_{\odot}$ and $\sim 1.5 M_{\odot}$) [45].

Table 1.1: Overview of the outstanding events seen during the third LVK observing run.

of GR and cosmology by comparing the information obtained in the two observation channels [39]. Along with opening new avenues to study such events in our Universe, this constitutes additional proof that the GW detections are genuine by independently observing related phenomena. It also marks the start of the multi-messenger astronomy (MMA) era, where several independent observation channels can be used to study a given physical process.

All the events mentioned up to now are part of the early stages of direct GW detection, being observed during the first and second observation runs, where each GW observation was exceptional on its own. Together, the runs spanned from September 2015 to August 2017, with a gap from January 2016 to November 2017 to upgrade the detectors. Over this period, 10 BBHs and 1 BNS were detected [35], enhancing the confidence in direct GW detection. Even more recently, the detection of GWs has become more common, with about 90 CBCs observed after the third observation run (extending from the 1st of April 2019 until the 27th of March 2020, with a one-month gap in October 2019 for detector upgrade) completed [40]. While many events were detected, a few were of particular interest since they led to the observation of novel effects. They are summarized in Table 1.1.

On top of the discoveries brought by the analysis of exceptional events, the amount of observations also makes the statistical analysis of the events possible. A representation of all the events detected by the LIGO-Virgo-KAGRA (LVK) collaboration⁸ is given in Fig. 1.4, where one sees the masses of the progenitors and the mass of the final object. Combining the information from all the observed CBC coalescences, one can study the population of compact binaries present in our Universe, probing their merger

⁸For the two first observation runs, KAGRA had not yet joined the collaboration. From the third observation run on, it is included in the network, “upgrading” the collaboration from LIGO-Virgo to LIGO-Virgo-KAGRA (LVK) collaboration.

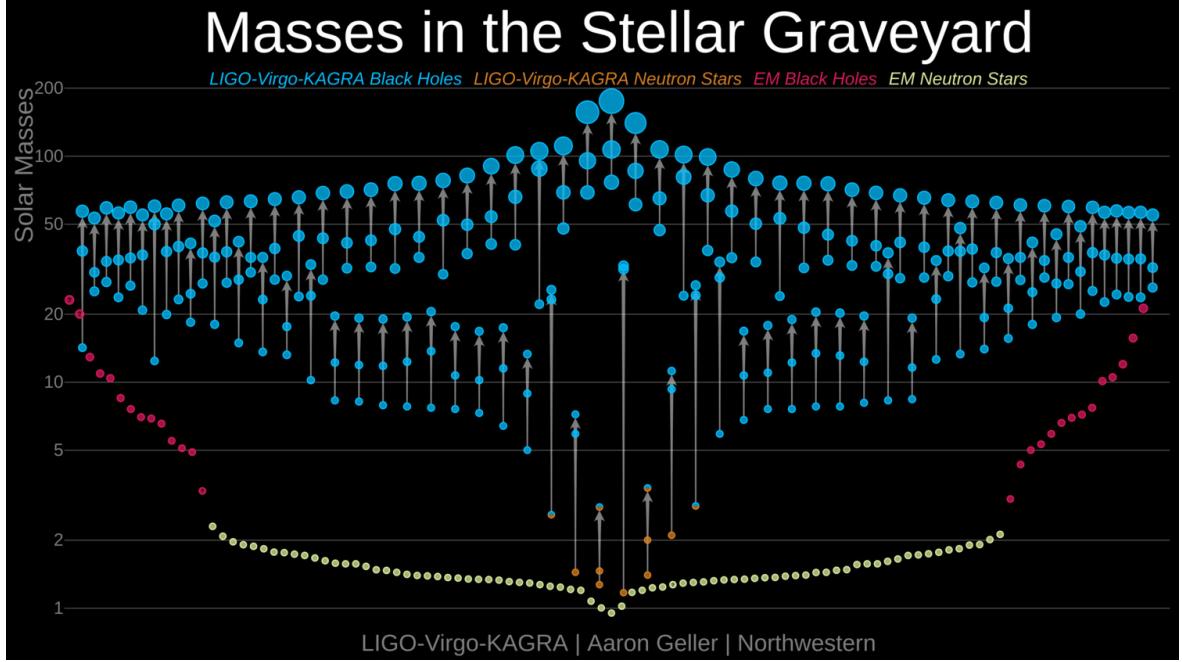


Figure 1.4: Representation of all the CBCs observed by the GW detection network and showing the initial masses of the components and the final mass of the resulting object. In blue and orange are the BHs and NSs observed by the LVK collaboration, while in red and yellow, one has the electromagnetically observed ones.

rate up to a redshift of 1, and constructing their mass and spin distributions [46]. In addition, one can stack the observed events to look for deviations from GR [47] and probe cosmology [48].

Whilst many breakthroughs have been achieved thanks to the current network of detectors, it is important to continue improving them to discover even more novelties. In that light, the detectors are regularly upgraded between observation runs to reach higher sensitivities. Current detectors, called second-generation (2G) detectors, are not yet at their maximal potential, and the next observing run should make the observation of CBC coalescences even more routine. In addition, since detectors have off-periods (due to noise for example), it is important to have more detectors to maximize the chances of having a large enough active detector network when interesting GW signals reach the Earth. Therefore, an additional LIGO detector, called LIGO-India is being built in the Dudhala village in the Hingoli district of Maharashtra [49]. Upgrades and additional detectors will lead to improved rates and probably the observation of new phenomena. One such avenue, treated in this thesis, is the gravitational lensing of GWs. However, there are fundamental limitations to the detectors, due to their size and design, making it impossible to observe some predicted science cases (for example higher mass BHs, the merger and post-merger parts of an NS signal) and probably also some unknown phenomena. Therefore, the next generation of detectors (called 3G detectors) is also

planned.

For the ground-based detector network, there is the European ET [30, 31] and the American Cosmic Explorer (CE) [50, 51]. While the currently planned design for the two detectors is quite different, they will both have much larger sizes than the current detectors, with ET planned to be a triangular detector with 10 km sides, and CE being 2 L-shaped detectors with arm lengths between 20 and 40 km. This, and other technological improvements, should push the reach of the detectors to the cosmic dawn for BBHs, making it able to probe periods where only primordial BHs can exist [52]. Thanks to improved low-frequency sensitivity, ET will also have the capacity to detect many more higher-mass systems, leading to greater opportunities to observe intermediate-mass BHs (IMBHs) [52]. A representation of the horizon redshift as a function of mass for ET and CE is given in Fig. 1.5.

Another planned detector is the Laser Interferometer Space Antenna (LISA) [53], a space-based detector with a triangular configuration trailing the Earth in its orbit around the Sun. While the ground-based detectors have inherent noises, limiting their sensitivity range from 1 Hz⁹ to thousands of Hz, LISA will focus on the range from 10^{-5} Hz to 1 Hz. Therefore, it will probe other types of objects. To cite only a few examples, it would see super-massive BBH mergers, extreme mass ratio inspirals, and white dwarf binaries far from merger. It has also the capacity to detect wider orbits for lower mass BBHs, opening the possibility to perform multi-band GW observations, by first tracking the evolution in the LISA band, and later following the merger in the 3G detectors [53].

To complete the current and future methods to probe GWs, pulsar timing arrays (PTAs) are also used to search for GWs (for example, see [54] for an overview). Here, millisecond pulsars are used as astronomical clocks. Owing to their stable rotation period, if a GW passes, it will modify the time of arrival of the EM pulses. Using different pulsars as a baseline and checking the differential modification, one has a much longer baseline, probing even lower frequencies than with LISA. In 2021, the NANOGrav Collaboration released its 12.5 years of data resulting from the observation of 47 pulsars [55]. This was used, for example, to constrain individual super-massive BHs [56], put constraints on the GW background (GWB) [57, 58], search for cosmological phase transition [59], and test GR by looking for additional polarizations not predicted by the theory [60]. One of the main recent results is the one on GWB as the first signs of the background have been detected, and confident detection is forecast within 5 years [57, 58]. A comparison of the different frequency ranges probed by the different methods is shown in Fig. 1.6. One sees that by combining all of them, we can probe frequencies from $\sim 10^{-10}$ Hz to $\sim 10^4$ Hz, covering a wide range of objects and processes.

Because GW astronomy is successful and developing, one may start wondering what science goals will be feasible. The increase in the number of detection, as well as the duration of the signals (larger frequency range means that the signal stays longer in-band), will put considerable stress on the analysis pipelines, requiring both enhancement

⁹This is the best-case scenario for the ET. LIGO detectors cannot go under 10 Hz.

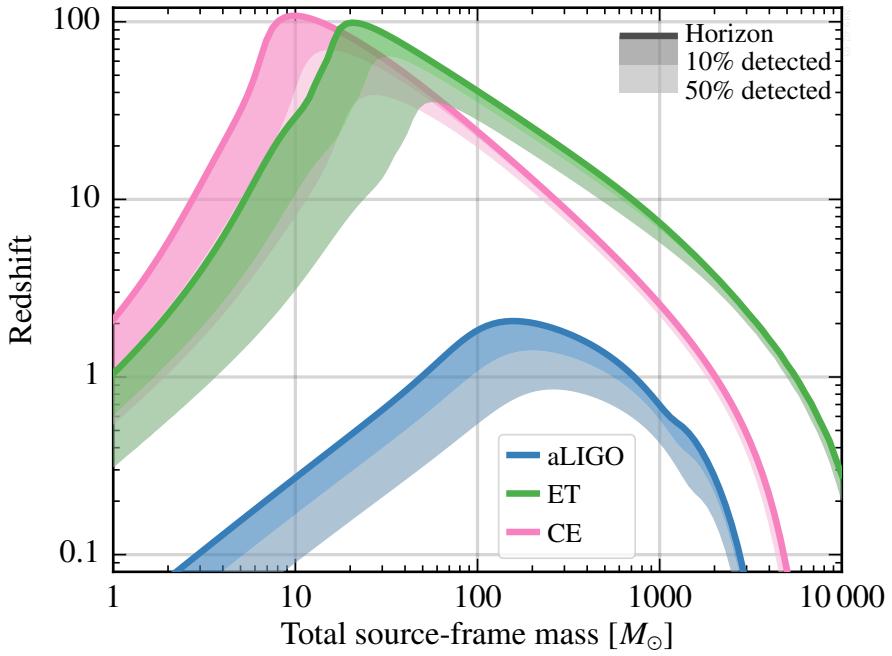


Figure 1.5: Horizon redshift for the different masses for the ET and the CE. The redshift extends up to cosmic dawn for some masses. Figure from [52].

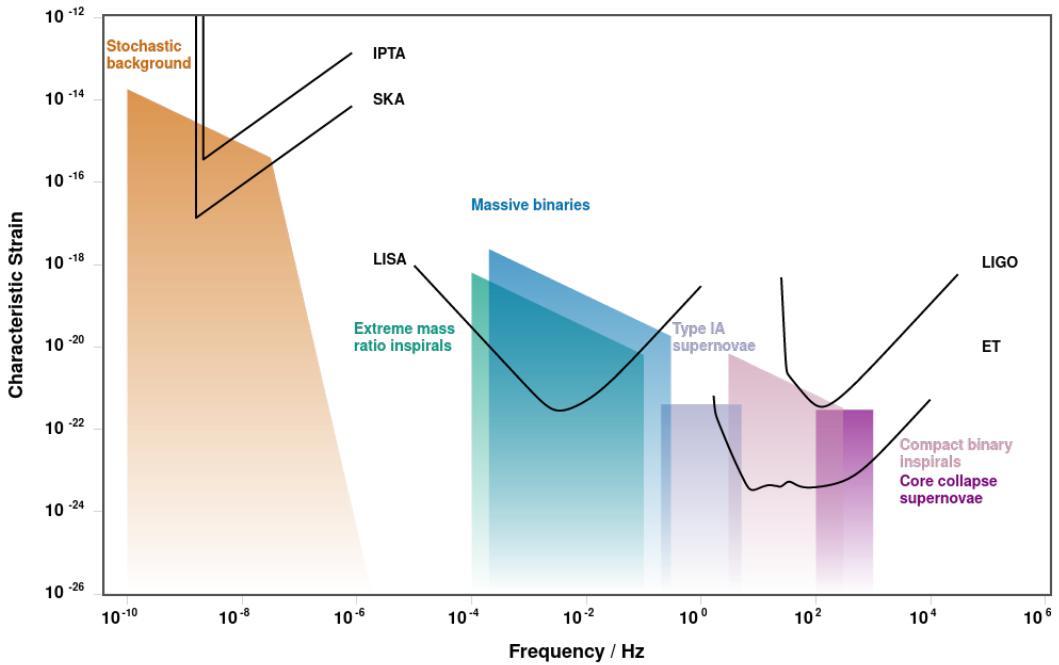


Figure 1.6: Overview of the sensitive frequency range for the different GW detection methods with examples of the objects that can be probed. Figure generated with GWPLOTTER [61].

of current methods and the introduction of new ones. Such methods would then still need to keep a high precision as having errors in the analysis step would lead to biases in the extracted information. To circumvent some of these issues, machine learning (ML)-based frameworks are growing in the context of GWs as their speed after training would be a major advantage in the future (see Ref. [62] for a review). Overall, though, progress still needs to be made on the theoretical and data analysis sides to use all of these instruments optimally.

In this thesis, several aspects of GW science are treated. We show the development of a data analysis tool for the detection of GW strong lensing. We also expose some of the issues one can face when analyzing 3G detector data and suggest some solutions. In Part. I, we give the various theoretical concepts required to understand the results presented in this thesis. More precisely, in Chapter 2, we give the main GR concepts needed to understand the generation and propagation of GWs and derive the latter. In Chapter 3, we discuss waveforms, a representation of GW signals for CBC mergers, and indispensable for their analysis. Then, in the next chapter, we discuss how one analyzes GW data. In Chapter 5, we explain lensing and how it should manifest itself in GW data. Chapter 6 introduces the various ML concepts used in this thesis. The three main topics studied in this thesis are then represented in the different parts. Part II explains the various lensing studies performed, going from the development of a new and more efficient analysis pipeline (Chapter 8) to an investigation of the science potential (Chapter 9), to the study of the false-alarm probability related to them (Chapter 10). Part. III shows results concerning the use of ML for GW early alerts. In Chapters 12 and 13, we show how ML identify incoming BNS signals before their merger, and in Chapter 14, we show how it can rapidly provide sky maps. In Part IV, we look at possible issues in the data analysis for 3G detectors. First, in Chapter 16, we start by establishing that overlapping signals will happen, before showing the biases they can lead to. Then, in Chapter 17, we propose Bayesian analysis methods to avoid biases. In Chapter 18, we show how ML approaches can be used to analyze overlapping signals. Finally, in Part. V, we give an overview of the thesis, and concluding remarks.

Part I

INTRODUCTION TO GRAVITATIONAL
WAVES, RELATED PHENOMENOLOGY,
AND OTHER USEFUL TOOLS

2

GENERAL RELATIVITY AND GRAVITATIONAL WAVES

In this chapter, we give a brief reminder on GR, focusing on the aspects needed to understand GW physics and the research carried out in this work. After reminding the reader of the main ingredients of GR, we turn to various scenarios outlining the emission, propagation, and detection of GWs.

2.1 The Main Ingredients of General Relativity

In GR, spacetime is described by a 4-dimensional manifold with a metric tensor $g_{\mu\nu}$ ¹ giving it its geometrical structure. The metric tensor can be used to define an invariant length element

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu, \quad (2.1)$$

where we use *Einstein's summation convention*². The metric tensor is also such that $g_{\mu\nu} g^{\nu\alpha} = \delta_\mu^\alpha$, and it can be used to lower and raise indices.

For a given situation, the square root of the line element will serve as a measure in spacetime. We define different types of interval depending on ds^2 : time-like intervals if $ds^2 > 0$, light-like intervals if $ds^2 = 0$, and space-like intervals if $ds^2 < 0$. Putting ourselves in the Minkowski metric, corresponding to a flat spacetime with the metric tensor $\eta = \text{diag}(-1, 1, 1, 1)$, the two first intervals are physically motivated as they correspond to an object moving at a speed smaller than the speed of light, or a photon (or other massless particles) moving at the speed of light. On the other hand, the last scenario would require the particle under consideration to move faster than the speed of light, which would break causality and is not physical.

In curved spacetime, the metric will also define a first-order derivation operator, or *a connection*, acting on the tensor field. It is called the *covariant derivative*, written ∇_α , and obeys the compatibility requirement with the metric

$$\nabla_\alpha g_{\mu\nu} = 0. \quad (2.2)$$

¹As is customary, we denote spacetime indices with Greek letters and space components with Roman letters.

²Each time an index is repeated in an “up” and “down” position, the summation is implied.

The covariant derivative is related to the coordinate derivatives $\frac{\partial}{\partial x^\alpha} = \partial_\alpha$ (where x^α represents the coordinate system) as

$$\nabla_\alpha u^\beta = \partial_\alpha u^\beta + \Gamma_{\alpha\delta}^\beta x^\delta, \quad (2.3)$$

with

$$\Gamma_{\gamma\delta}^\beta = \frac{1}{2} g^{\beta\rho} (\partial_\delta g_{\gamma\rho} + \partial_\gamma g_{\delta\rho} - \partial_\rho g_{\gamma\delta}), \quad (2.4)$$

the *Christoffel symbols*. These are symmetric in their lower indices.

The covariant derivative leads to the notion of parallel transport, describing how a field changes along a curve. If \vec{U} is the vector tangent to a curve \mathcal{C} , then a field \mathbf{T} is parallel transported along \mathcal{C} if

$$\nabla_{\vec{U}} \mathbf{T} = 0, \quad (2.5)$$

with $\nabla_{\vec{U}} = U^\alpha \nabla_\alpha$.

When a curve parallel transports its tangent vector, it is called a *geodesic*. So a curve $\mathcal{C}(\lambda)$ with a tangent vector \vec{U} is a geodesic if

$$\nabla_{\vec{U}} \vec{U} = 0, \quad (2.6)$$

which in terms of components and coordinates is

$$\frac{d^2 x^\mu}{d\lambda^2} + \Gamma_{\alpha\beta}^\mu \frac{dx^\alpha}{d\lambda} \frac{dx^\beta}{d\lambda} = 0. \quad (2.7)$$

This is a second order differential equation for x^μ and has a unique solution when the initial conditions $x^\mu(\lambda_0)$ and $(dx^\mu/d\lambda)(\lambda_0)$ are specified. For a flat space, this reduces to a straight line. From a physical perspective, geodesics are important as time-like geodesics are the spacetime trajectories of free-falling particles.

In GR, the motion of particles depends on the structure of spacetime. Its geometry is related to the matter distribution within it through the *Einstein Field Equations* (EFEs):

$$G_{\alpha\beta} = -\frac{8\pi G}{c^4} T_{\alpha\beta}, \quad (2.8)$$

where c is the usual constant speed of light, and G is the gravitational constant. On the left-hand side, $G_{\alpha\beta}$ is the *Einstein Tensor* representing the spacetime geometry. It is made of the metric tensor and derivatives of the Christoffel symbols. It vanishes when considering a flat spacetime. On the right-hand side, $T_{\alpha\beta}$ is the stress-energy tensor, respecting the conservation law $\nabla_\alpha T^{\alpha\beta} = 0$ and representing the effect of matter. The equations can be derived from principles similar to those used in Newtonian gravity, where the Einstein tensor plays the role of the field, and the stress-energy tensor plays the role of the density. The EFEs are such that they reduce to the usual Newtonian Poisson equation in the weak-field limit. Eq (2.8) represents ten coupled second-order non-linear partial differential equations governing the behavior of matter and spacetime.

They generally do not have analytical solutions, except in a few cases, where one can use symmetries to simplify the problem.

2.2 First Hints of Gravitational Waves

Based on the ingredients presented in the previous section, we can derive the basics of GW generation, propagation, and detection. Therefore, we start by looking into *linearized* GR. This will have solutions that can be viewed as waves in spacetime.

2.2.1 Linearized General Relativity

Because gravitation is a weak effect in the absence of dense objects, it is often possible to describe the metric tensor as the sum of a background metric ($\tilde{\eta}$) and a small perturbation (h). If we are far away from any source, the metric is decomposed as

$$g_{\alpha\beta} = \tilde{\eta}_{\alpha\beta} + h_{\alpha\beta}, \quad (2.9)$$

where $\|h\| \ll 1$, and we neglect effects of the order $\mathcal{O}(h^2)$. Therefore, $\tilde{\eta}_{\alpha\beta}$ can be used to raise on lower indices, and we have

$$g^{\alpha\beta} = \tilde{\eta}^{\alpha\beta} - h^{\alpha\beta}, \quad (2.10)$$

from applying $\tilde{\eta}$ on Eq. (2.9), and noting $h^{\alpha\beta} = \tilde{\eta}^{\alpha\gamma}\tilde{\eta}^{\beta\delta}h_{\gamma\delta}$.

In this regime, if $\tilde{\eta} = \eta$, the Minkowski tensor, we look at the evolution of the tensor field h on a flat background spacetime. This is the situation we consider here. Then, defining the trace $h = h^\alpha_\alpha$ and the d'Alembertian operator $\square = \partial^\alpha_\alpha$, the Einstein tensor, at first order in h , reduces to

$$G_{\alpha\beta} = \frac{1}{2} \left(\partial_\gamma \partial_\beta h^\gamma_\alpha + \partial_\gamma \partial_\alpha h^\gamma_\beta + \eta_{\alpha\beta} \square h - \partial_\alpha \partial_\beta h - \square h_{\alpha\beta} - \eta_{\alpha\beta} \partial_\gamma \partial_\delta h^{\gamma\delta} \right), \quad (2.11)$$

which can directly be used in Eq. (2.8) to give the linearized EFEs at the first order in h .

However, we can make the expression easier to use by performing a *gauge transformation*. Under the assumption that $h_{\alpha\beta}$ is a small perturbation, we are allowed to make an infinitesimal change in coordinates

$$x'^\alpha = x^\alpha + \xi^\alpha(x^\beta), \quad (2.12)$$

provided $\partial_\beta \xi^\alpha \leq |h_{\alpha\beta}|$. Indeed, under this coordinate transform, the perturbation

changes as

$$h'_{\alpha\beta} = h_{\alpha\beta} - (\partial_\alpha \xi_\beta + \partial_\beta \xi_\alpha), \quad (2.13)$$

showing that we need the $\partial_\beta \xi^\alpha \leq |h_{\alpha\beta}|$ condition to be respected. Such a transformation is called a gauge transformation.

In parallel to the local one, global transformations, such as the Lorentz transformation, can be applied:

$$x'^\alpha = \Lambda_\beta^\alpha x^\beta. \quad (2.14)$$

Using the Lorentz transformation and the gauge choice, we can compute a simplified version of the EFEs. Therefore, we first define and work with the *trace-reversed* perturbation tensor \bar{h} . By definition, $\bar{h}_\alpha^\alpha = -h$ so that

$$\bar{h}^{\alpha\beta} = h^{\alpha\beta} - \frac{1}{2} \eta^{\alpha\beta} \bar{h}. \quad (2.15)$$

Starting from an arbitrary \bar{h}^{old} perturbation such that Eq. (2.9) is respected, we can apply the gauge transformation from (2.12), and take the divergence

$$\partial_\beta \bar{h}^{\text{new},\alpha\beta} = \partial_\beta \bar{h}^{\text{old},\alpha\beta} - \partial_\beta \xi^{\alpha\beta}. \quad (2.16)$$

If we want a gauge where $\partial_\beta \bar{h}^{\text{new},\alpha\beta} = 0$, we need to impose the condition $\square \xi^\alpha = \partial_\beta \bar{h}^{\text{old},\alpha\beta}$. This condition is an inhomogeneous wave equation and always has a solution. Therefore, we can always find ξ^α leading to a gauge where $\partial_\beta \bar{h}^{\alpha\beta} = 0$. Such a gauge is called the *Lorenz* (or *de Donder*) gauge.

Placing ourselves in the Lorenz gauge, the linearized EFEs (2.11) take the form

$$\square \bar{h}_{\alpha\beta} = -\frac{16\pi G}{c^4} T_{\alpha\beta}. \quad (2.17)$$

Even if this expression is simplified, we can place ourselves in an even more advantageous gauge. Indeed, the condition $\partial_\alpha \bar{h}^{\beta\alpha} = 0$ gives four constraints, and $\bar{h}^{\beta\alpha}$ has ten possible components³. Therefore, one has 6 degrees of freedom left and can use them to simplify the problem a bit further. One can impose $\square \xi^\alpha = 0$, which gives 4 more constraints. This leaves only 2 degrees of freedom for \bar{h} . These are physical degrees of freedom and cannot be further simplified. So, there are only two independent components for $\bar{h}_{\alpha\beta}$. As we will see, this will lead to two polarizations for GWs. If we are far from any source of gravitation, one can fix these two remaining degrees of freedom by demanding (i) $\bar{h}_{0\alpha} = 0$, and (ii) imposing a traceless property $\bar{h}_\alpha^\alpha = 0$. Defining $u = (c, 0, 0, 0)$, the 4-velocity of a time-like observer, (i) can be recast as $u^\alpha \bar{h}_{\alpha\beta} = 0$, meaning that information will propagate in a direction perpendicular to the worldline of the observer. Imposing the conditions presented above brings us to the *transverse traceless gauge* (TT-gauge).

³Strictly speaking it should be 16. This reduces to 10 independent components when accounting for symmetry.

2.2.2 Let There be Waves

Far away from any source, we may assume vacuum ($T_{\alpha\beta} = 0$). In this case, Eq.(2.17) becomes

$$\square \bar{h}_{\alpha\beta} = 0, \quad (2.18)$$

which corresponds to a wave equation. It admits a general solution of the form

$$\bar{h}_{\alpha\beta} = \mathbb{R} \left(C_{\alpha\beta} e^{ik_\gamma x^\gamma} \right). \quad (2.19)$$

In this expression, $C_{\alpha\beta}$ represents the components of a tensor, while k_γ are the components of a 4-vector. To satisfy the wave-equation, this 4-vector needs to be $\mathbf{k} = (\omega, \vec{k})$, with $\omega = c|\vec{k}|$.

Using the Lorenz gauge condition, $k^\alpha C_{\alpha\beta} = 0$, and using the traceless gauge, $C_\alpha^\alpha = 0$, implies that $C_{0\alpha} = 0$.

Choosing a coordinate system such that the z-axis points in the wave's direction of propagation, the wave vector \mathbf{k} takes the form $(\omega, 0, 0, \omega)$. The Lorenz gauge condition then implies $C_{\alpha z} = 0$, and we are left only with the x and y components of C . Since we are also working in the traceless gauge, $C_{xx} = -C_{yy}$, and because the tensor is symmetric, $C_{xy} = C_{yx}$. Therefore, we end up with two components with a physically-important meaning: C_{xx} and C_{xy} . Explicitly, the plane-wave solution in the TT gauge is

$$\bar{h}^{\text{TT}}(t, z) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & C_{xx} & C_{xy} & 0 \\ 0 & C_{xy} & -C_{xx} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cos(\omega(t - z/c)). \quad (2.20)$$

To understand how such a wave interacts with matter, one needs the expression for the remaining components.

2.2.3 Basic Effects of Gravitational Waves on Matter

Considering a plane-wave propagating along the z axis, the metric in presence of a GW is $g_{\alpha\beta} = \eta_{\alpha\beta} + \bar{h}_{\alpha\beta}^{\text{TT}}$, and the line element (see Eq. (2.1)) reduces to

$$\begin{aligned} ds^2 = & -c^2 dt^2 + \left[1 + C_{xx} \cos(\omega(t - z/c)) \right] dx^2 + \left[1 - C_{xx} \cos(\omega(t - z/c)) \right] dy^2 \\ & + 2C_{xy} \cos(\omega(t - z/c)) dx dy + dz^2. \end{aligned} \quad (2.21)$$

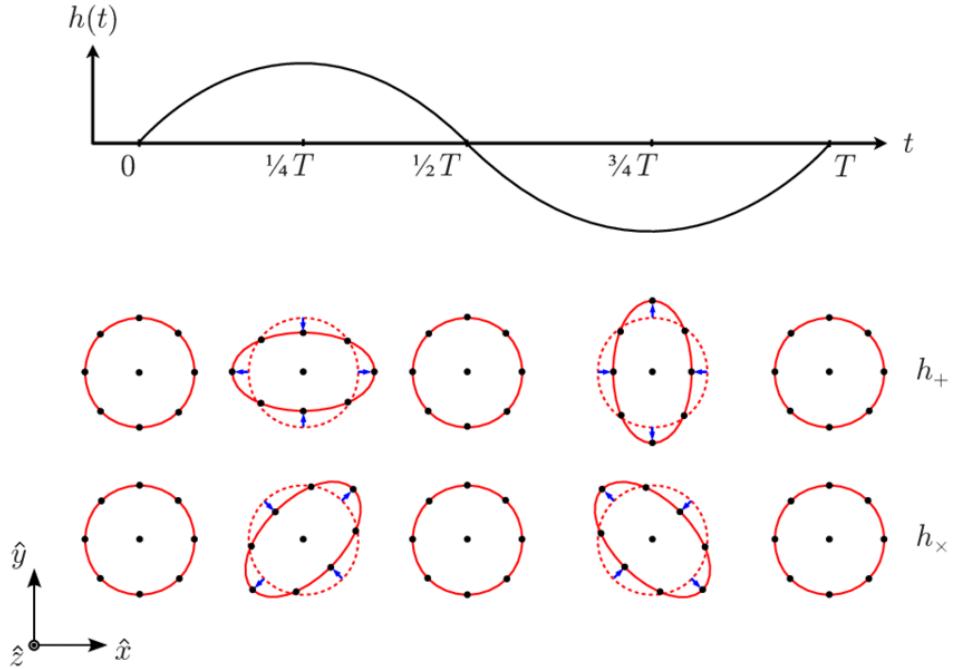


Figure 2.1: Representation of the plus (2nd line) and cross-polarization (3rd line) of a GW, with their evolution over a period. The GW is assumed to propagate in the \hat{z} direction. For this representation, one sets the complementary polarization to zero. However, in reality, the GW is made of the superposition of the two polarizations. Figure adapted from [63].

Taking $C_{xy} = 0$ and $C_{xx} \neq 0$, the physical distances in the x and y direction stretch and squeeze periodically, in an opposite way, for the two directions. If one takes a ring particle, it means that it becomes elongated in the x direction at one time, and half a period later, it is elongated in the y direction. So, the deformation has a $+$ -shape, and one writes $C_{xx} = h_+$. This periodic deformation is represented in the middle row of Fig. 2.1, where the GW propagates along the z axis. On the other hand, if one takes $C_{xx} = 0$ and $C_{xy} \neq 0$, the deformation happens the same way but rotated by 45° ⁴. Therefore, the same sort of periodic squeezing and stretching happens for the circle of particles. This time it has a \times -shape in the (x, y) coordinates, and one writes $C_{xy} = h_\times$. This scenario is depicted in the lower panel of Fig. 2.1.

⁴This can be shown by mapping the (x, y) to (x', y') coordinates using a 45° rotation around the z axis.

2.2.4 Premises of Detection

The deformation induced by a passing GW already suggests how to detect them. Indeed, if instead of a ring of particles, one takes an L-shaped interferometer with its arms aligned along the x-axis and a test mass at their extremity, one can measure the $+$ -polarization through the arms' squeezing and stretching. Similarly, a detector rotated by 45° would detect the \times -polarization. If the interferometer has an arbitrary position, it would detect a mixture of the two polarizations.

We assume an L-shaped detector with arm-lengths L , aligned along the x and y axes. For LIGO and Virgo, as mentioned in Chapter 1, these lengths would be 4 and 3 km, respectively. Using Eq. (2.21), one can show that the test masses fixed at the end of the detectors would have a displacement

$$\delta L_x(t) = h_+(t) \frac{L}{2} \quad (2.22)$$

$$\delta L_y(t) = -h_+(t) \frac{L}{2}. \quad (2.23)$$

The measured strain at the output of the detector is the relative difference in arm length

$$h(t) = \frac{\delta L_x(t) - \delta L_y(t)}{L}, \quad (2.24)$$

which for this particular detector set up is $h_+(t)$. If we rotate the detector by 45° , we would instead measure $h_\times(t)$.

More realistically, the detector will have an arbitrary orientation compared to the incoming GW. Therefore, we need to translate the change in distances induced by the wave from the GW-frame (x', y', z') to the detector-frame (x, y, z) coordinates. Placing the detector arms along x and y, and the GW direction of propagation in the z' -direction, one can find the projection factors F_+ and F_\times such that

$$h(t) = F_+ h_+(t) + F_\times h_\times(t), \quad (2.25)$$

where we have not yet specified the dependencies for $F_{+,\times}$. It is easy to see that the angles contributing to the projection are the *sky location* of the source (θ, ϕ) , which describes how the origin of the axes are situated the one with respect to the other, and an angle that translates the extra rotation required to match the axes. This angle is the *polarization* angle ψ . A representation of the different angles is given in Fig. 2.2. Using each of these angles to perform a rotation enables one to go from one frame to the other.

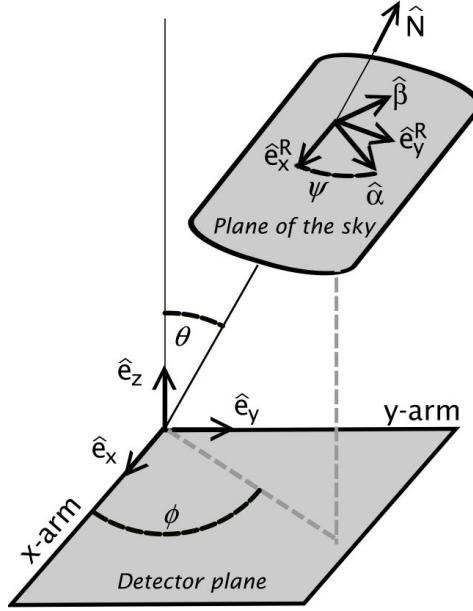


Figure 2.2: Representation of the angles linking the detector-frame axes $(\hat{e}_x, \hat{e}_y, \hat{e}_z)$ and the GW-frame axes $(\hat{e}_x^R, \hat{e}_y^R, \hat{e}_z^R)$. Adapted from [64].

Applying these three successive rotations to $h_{+, \times}(t)$, we get:

$$F_+(\theta, \phi, \psi) = \left(\frac{1 + \cos^2(\theta)}{2} \right) \cos(2\phi) \cos(2\psi) - \cos(\theta) \sin(2\phi) \sin(2\psi) \quad (2.26)$$

$$F_\times(\theta, \phi, \psi) = \left(\frac{1 + \cos^2(\theta)}{2} \right) \cos(2\phi) \sin(2\psi) + \cos(\theta) \sin(2\phi) \cos(2\psi), \quad (2.27)$$

the *antenna pattern response functions* for an L-shaped detector.

Other designs have been envisaged for future detectors. For example, ET is proposed to be made of three V-shaped detectors with an opening angle $\alpha = 60^\circ$. There is a direct relationship between its antenna patterns and those for an L-shaped interferometer, with $F_{+, \times}^V(\theta, \phi, \psi) = \sin(\alpha) F_{+, \times}(\theta, \phi, \psi)$.

2.3 Generation of Gravitational Waves

We now turn to the description of the generation of GWs. Starting from Eq. (2.17), the properties of the d'Alembertian mean this equation has the generic solution

$$\bar{h}_{\alpha\beta}(x) = -\frac{16\pi G}{c^4} \int d^4x' G(x - x') T_{\alpha\beta}(x'), \quad (2.28)$$

where x represents the measurement's place, and x' is at the source. $G(x - x')$ is a Green's function. By imposing boundary conditions, we can select a specific function. Imposing that no radiation is incoming from infinity, one gets the *retarded Green's function*:

$$G(x - x') = -\frac{1}{4\pi|\vec{x} - \vec{x}'|} \delta(t_{\text{ret}}^0 - x'^0), \quad (2.29)$$

where we have defined the *retarded time* $t_{\text{ret}}^0 = t - |\vec{x} - \vec{x}'|/c$.

Plugging this into Eq. (2.28) yields

$$\bar{h}_{ij}(x) = \frac{4G}{c^4} \int d^3 \vec{x}' \frac{T_{ij}(t_{\text{ret}}^0, \vec{x}')} {|\vec{x} - \vec{x}'|}. \quad (2.30)$$

In Eq. (2.30), the temporal components $T_{0\alpha}$ of the stress-energy tensor are not present anymore. While this can look like an issue at first glance, it is not since the conservation law between energy and momentum

$$\partial_\alpha T^{\alpha\beta} = 0 \quad (2.31)$$

links the temporal and spatial components of the tensor.

We can also project the solution found on the TT gauge using the projector

$$\Lambda_{ijkl} = P_{ik}P_{jl} - \frac{1}{2}P_{ij}P_{kl}, \quad (2.32)$$

where $P_{ij} = \delta_{ij} - n_i n_j$, and n_i is the normal in the direction of propagation of the wave.

Using this projector enables one to get the components in the TT gauge

$$h_{ij}^{\text{TT}}(x) = \frac{4G}{c^4} \Lambda_{ijkl} \int d^3 \vec{x}' \frac{T^{kl}(t_{\text{ret}}^0, \vec{x}')} {|\vec{x} - \vec{x}'|}. \quad (2.33)$$

This expression is useful to derive more intuitive solutions to the linearized EFEs.

2.3.1 Multipole Expansion

To know what the GW generated by a fluctuating stress-energy tensor looks like, one needs to solve Eq. (2.33). In all generality, this equation cannot be solved analytically. However, when far away from the source *i.e.* at a distance r much larger than the characteristic size of the source, one can approximate $|\vec{x} - \vec{x}'|$ by r . Then, Eq. (2.33) becomes

$$h_{ij}^{\text{TT}}(x) = \frac{4G}{c^4 r} \Lambda_{ijkl} \int d^3 \vec{x}' T^{kl}(t_{\text{ret}}^0, \vec{x}'), \quad (2.34)$$

where the retarded time is now $t_{\text{ret}} = t - r/c$.

Using the conservation law $\partial_\alpha T^{\alpha\beta} = 0$, one can show that

$$\int d^3\vec{x} T^{ij} = \frac{1}{c^2} \int d^3\vec{x} \ddot{T}^{00} x^i x^j - \int d^3\vec{x} T^{ij}, \quad (2.35)$$

where the overdot represent the time derivative. Then, defining the *mass quadrupole moment*

$$M^{ij} = \frac{1}{c^2} \int d^3\vec{x} T^{00} x^i x^j, \quad (2.36)$$

the integral of the stress-energy tensor (2.35) can be re-expressed as

$$\int d^3\vec{x} T^{ij} = \frac{1}{2} \ddot{M}^{ij}, \quad (2.37)$$

and the observed GW becomes

$$h_{ij}^{\text{TT}}(x) = \frac{2G}{c^4 r} \Lambda_{ijkl} \ddot{M}^{kl}(t_{\text{ret}}). \quad (2.38)$$

This expression is the leading-order effect of the GW emission. Alternatively, it can be recovered by Taylor expanding the stress-energy tensor and keeping the leading order.

As was done in the previous section, without loss of generality, we choose the coordinates so that the z-axis points toward the observer and Eq. (2.38) becomes

$$h_{ij}^{\text{TT}} = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (2.39)$$

Choosing the normal along z means $P_{11} = P_{22} = 1$ and $P_{12} = P_{21} = P_{j3} = 0$ in Eq. (2.32). Then, defining

$$A_{ij} = \Lambda_{ijkl} \ddot{M}^{kl}, \quad (2.40)$$

one can show that $A_{j3} = 0$, and

$$A_{11} = -A_{22} = \frac{\ddot{M}^{11} - \ddot{M}^{22}}{2} \quad (2.41)$$

$$A_{21} = A_{12} = \ddot{M}^{12}. \quad (2.42)$$

Completing this with the prefactor from Eq. (2.34) and comparing with the elements from Eq. (2.39), one finds

$$h_+ = \frac{G}{rc^4} (\ddot{M}^{11} - \ddot{M}^{22}) \quad (2.43)$$

$$h_\times = \frac{2G}{rc^4} \ddot{M}^{12}, \quad (2.44)$$

which is a generic form for the GW emission at leading order.

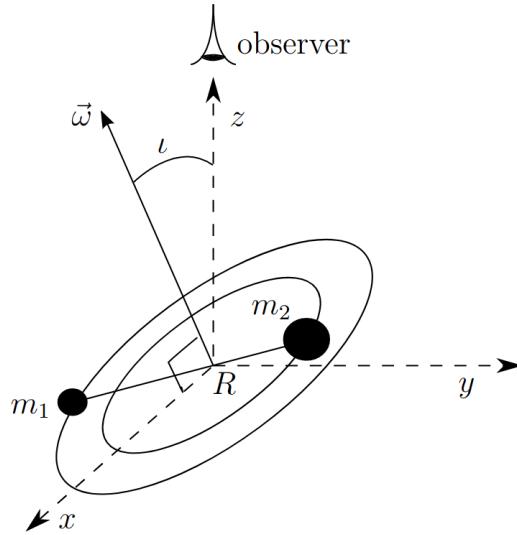


Figure 2.3: Representation of a binary system made of two point-mass particles on a fixed circular orbit. The coordinates are such that the z -axis points to the observer and the angle between the plane of the orbit and the line of sight is ι . The origin is placed at the center of mass of the system. The two masses are at a fixed distance R and have an orbital frequency ω . Figure adapted from [65].

2.4 Gravitational-Wave Emission from Rotating Objects

As a prelude to more complex emission schemes, let us consider two point-mass particles m_1 and m_2 moving along a circular orbit⁵. The particles orbit each other at a distance R with orbital frequency ω . We work in coordinates where the z -axis points towards the observer, and define the *inclination angle* ι – the angle between the plane of the orbit and the line of sight. Fig. 2.3 gives a representation of the situation.

If we also choose the coordinates so that the origin is aligned with the center of mass of the system, the position of the particles $\vec{x}_1(t)$ and $\vec{x}_2(t)$ are

$$\vec{x}_1(t) = \frac{\mu R}{m_1} \hat{e}(t) \quad (2.45)$$

$$\vec{x}_2(t) = -\frac{\mu R}{m_2} \hat{e}(t), \quad (2.46)$$

where $\hat{e}(t) = (\cos(\omega t), \cos(\iota) \sin(\omega t), \sin(\iota) \sin(\omega t))$ is the unit vector pointing from m_2 to m_1 for our particular choice of coordinates, and $\mu = m_1 m_2 / (m_1 + m_2)$ is the reduced

⁵We momentarily neglect the energy carried away by GWs.

mass.

As an approximation, we consider only the quadrupole emission (2.36). If the two masses are still far away and non-relativistic, the quadrupole moment is

$$\begin{aligned} M^{ij}(t) &= \frac{1}{c^2} \int d^3\vec{x} T^{00}(t, \vec{x}) x^i x^j = \int d^3\vec{x} \rho(t, \vec{x}) x^i x^j \\ &= \int d^3\vec{x} \left[m_1 \delta\left(\vec{x} - \frac{\mu R}{m_1} \hat{e}(t)\right) + m_2 \delta\left(\vec{x} + \frac{\mu R}{m_2} \hat{e}(t)\right) \right] = \mu R e^i e^j, \end{aligned} \quad (2.47)$$

where ρ is the mass density, expressed in terms of a Dirac delta function using the point mass particle's coordinate.

Performing the time derivative of \ddot{M}^{11} , \ddot{M}^{12} and \ddot{M}^{22} , plugging them into Eq. (2.43) and evaluating the quadrupole at the restarted time, the polarizations of the emitted GW are

$$\begin{aligned} h_+(t) &= -\frac{4\mu G R^2 \omega^2}{r c^4} \left(\frac{1 + \cos^2(\iota)}{2} \right) \cos(2\omega t_{\text{ret}}), \\ h_\times(t) &= -\frac{4\mu G R^2 \omega^2}{r c^4} \cos(\iota) \sin(2\omega t_{\text{ret}}). \end{aligned} \quad (2.48)$$

Under the approximation that the masses are far from each other, Kepler's third law

$$\omega^2 = \frac{GM}{R^3}, \quad (2.49)$$

can be used to relate ω , R and the masses. Also defining the *chirp mass*

$$\mathcal{M}_c = \frac{(m_1 m_2)^{\frac{3}{5}}}{(m_1 + m_2)^{\frac{1}{5}}}, \quad (2.50)$$

the polarizations take the form

$$\begin{aligned} h_+(t) &= -\frac{4}{r} \left(\frac{G \mathcal{M}_c}{c^2} \right)^{\frac{5}{3}} \left(\frac{\omega}{c} \right)^{\frac{2}{3}} \left(\frac{1 + \cos^2(\iota)}{2} \right) \cos(2\omega t_{\text{ret}}), \\ h_\times(t) &= -\frac{4}{r} \left(\frac{G \mathcal{M}_c}{c^2} \right)^{\frac{5}{3}} \left(\frac{\omega}{c} \right)^{\frac{2}{3}} \cos(\iota) \sin(2\omega t_{\text{ret}}). \end{aligned} \quad (2.51)$$

These equations show that the GWs are emitted with twice the orbital frequency of the two test masses, and the period of the GW is half that of the orbiting binary. The strength of the observed polarization depends on the inclination angle. If $\iota = 90^\circ$, we see the binary from the edge, and only the +-polarization is observable. On the other hand, when the plane of the binary is perpendicular to the line of sight, the two polarizations have the same contribution.

2.5 Inspiraling Objects

While the derivation done in the previous section already brings important information about GW emission from objects rotating around each other, it neglects many effects. In particular, it disregards the energy carried away by the GW, which should lead to a shrinking distance between the particles and an increased orbital frequency.

As an illustrative example, we can consider the case of two point masses. Defining $\Phi(t)$ the time-dependent angle swept by the line connecting the two point masses over time, we have

$$\omega(t) = \dot{\Phi}(t), \quad (2.52)$$

and the positions of the two particles over time

$$\begin{aligned} \vec{x}_1(t) &= \frac{\mu R(t)}{m_1} \hat{e}(t), \\ \vec{x}_2(t) &= -\frac{\mu R(t)}{m_2} \hat{e}(t), \end{aligned} \quad (2.53)$$

where $\hat{e}(t) = (\cos(\Phi(t)), \cos(\iota) \sin(\Phi(t)), \sin(\iota) \sin(\Phi(t)))$ is the unit vector pointing from m_2 to m_1 , which now accounts for the change in frequency over time.

We assume the orbit to be slowly decaying. So, for each revolution, the trajectory can be approximated by a circle. Using this adiabatic assumption, Eq. (2.51) becomes

$$\begin{aligned} h_+(t) &= -\frac{4}{r} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{\frac{5}{3}} \left(\frac{\omega(t_{\text{ret}})}{c} \right)^{\frac{2}{3}} \left(\frac{1 + \cos^2(\iota)}{2} \right) \cos(2\Phi(t_{\text{ret}})), \\ h_\times(t) &= -\frac{4}{r} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{\frac{5}{3}} \left(\frac{\omega(t_{\text{ret}})}{c} \right)^{\frac{2}{3}} \cos(\iota) \sin(2\Phi(t_{\text{ret}})), \end{aligned} \quad (2.54)$$

where we account for the time-dependency of the different quantities.

We know that the system is losing energy through the radiation of GWs. The origin of this energy is the system's orbital energy (E_{orb}). Therefore,

$$\frac{dE_{\text{orb}}}{dt} = -\frac{dE}{dt}, \quad (2.55)$$

where E is the energy of the GW.

Assuming non-relativistic particles so that we may use the Newtonian formalism,

$$E_{\text{orb}} = \frac{1}{2}m_1v^2 + \frac{1}{2}m_2v^2 - \frac{Gm_1m_2}{R}. \quad (2.56)$$

Using Kepler's third law (2.49) and

$$v_i = \frac{\mu R \omega_i}{m_i}, \quad (2.57)$$

where $i = \{1, 2\}$, we find

$$E_{\text{orb}} = -\frac{1}{2}(G^2 \mathcal{M}_c^5 \omega^2)^{\frac{1}{3}}, \quad (2.58)$$

where the dependence on the orbit's size has been eliminated.

For the GW power emitted, one can show that in the TT gauge, it takes the form [24]

$$\begin{aligned} \frac{dE}{dt} &= \frac{c^3 r^2}{16\pi G} \int d\Omega \langle \dot{h}_+^2 + \dot{h}_\times^2 \rangle_t \\ &= \frac{32c^5}{5G} \left(\frac{G\mathcal{M}_c\omega}{c^3} \right)^{\frac{10}{3}}, \end{aligned} \quad (2.59)$$

where we have used Eq. (2.54) for $h_{+,\times}(t)$.

Defining the GW frequency $f_{\text{gw}} = 2f_{\text{orb}}$, and using Eqs (2.58) and (2.59), Eq. (2.55) becomes a differential equation for the frequency

$$\dot{f}_{\text{gw}}(t_{\text{ret}}) = \frac{96\pi^{\frac{8}{3}}}{5} \left(\frac{G\mathcal{M}_c}{c^3} \right)^{\frac{5}{3}} f_{\text{gw}}^{\frac{11}{3}}(t_{\text{ret}}), \quad (2.60)$$

with the solution

$$f_{\text{gw}}(t) = \frac{1}{\pi} \left(\frac{G\mathcal{M}_c}{c^3} \right)^{-\frac{5}{8}} \left(\frac{5}{256(t_c - t)} \right)^{\frac{3}{8}}. \quad (2.61)$$

t_c is the *coalescence time* and is a constant of integration. In particular, $f_{\text{gw}} \rightarrow \infty$ when $t \rightarrow t_c$. However, this would imply a null distance between the two objects. In reality, it is more representative of the moment when the simplifying approximation breaks. Real mergers are systems containing finite-sized objects like BHs and NSs, and the inspiral would terminate before the distance is zero. It could be when the objects touch, but in fact, it can be even earlier than that. This can be shown only by using a full GR derivation; here we give a rough argument. Eq. (2.61) shows an inspiraling motion such that the frequency increases, and it does so faster and faster, leading to a runaway process, seemingly leading to this infinite frequency. However, using Schwarzschild geometry, one can compute a distance, R_{ISCO} , corresponding to the innermost stable orbit of the particles, hence the smallest stable orbit possible. In fact [24]

$$R_{\text{ISCO}} \simeq \frac{6GM}{c^2}, \quad (2.62)$$

where $M = m_1 + m_2$ is the total mass of the system. This represents the distance at which the quasi-circular orbit approximation breaks down. Once reached, the two

objects plunge toward each other and merge.

Inserting Eq (2.62) in Kepler's third law (2.49), we find the GW frequency for which the approximation breaks down [24]

$$f_{\text{ISCO}} = \frac{c^3}{6^{\frac{3}{2}} \pi G M} . \quad (2.63)$$

Defining $\Phi_{\text{gw}} = 2\Phi$, and integrating Eq. (2.61) yields

$$\Phi_{\text{gw}}(t) = -2 \left(\frac{5G\mathcal{M}_c}{c^3} \right)^{-\frac{5}{8}} (t_c - t)^{\frac{5}{8}} + \Phi_c , \quad (2.64)$$

where Φ_c is a constant of integration corresponding to the phase when $t_c = t$.

Using Eq. (2.64), the final form of the GW polarization is

$$\begin{aligned} h_+(t) &= -\frac{4}{r} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{\frac{5}{3}} \left(\frac{\pi f_{\text{gw}}(t_{\text{ret}})}{c} \right)^{\frac{2}{3}} \left(\frac{1 + \cos^2(\iota)}{2} \right) \cos(\Phi_{\text{gw}}(t_{\text{ret}})) , \\ h_{\times}(t) &= -\frac{4}{r} \left(\frac{G\mathcal{M}_c}{c^2} \right)^{\frac{5}{3}} \left(\frac{\pi f_{\text{gw}}(t_{\text{ret}})}{c} \right)^{\frac{2}{3}} \cos(\iota) \sin(\Phi_{\text{gw}}(t_{\text{ret}})) . \end{aligned} \quad (2.65)$$

The change in the polarizations over time will also change the power emitted by the GW and then its dynamics. However, GW inspirals due to CBC mergers in GR have qualitatively the same behavior.

3

WAVEFORMS: KEY TO ANALYSIS OF THE DATA

In Chapter 2, we have seen how from the basic components of GR, one can arrive at inspiraling point masses. However, it has also been explained that the approximations used have their limits and cannot describe the GW behavior further than the ISCO frequency (2.63). In addition, the main objects studied throughout GW emissions are CBCs made from two dense bodies circling each other, losing energy through GW emission before merging. In particular, at the time of writing, detected GW signals correspond to BBH, BNS, and NSBH mergers [40]. Their inspiral is qualitatively the same as the one described by Eq. (2.65), but a more complete approach is needed to describe the system from inspiral to after its merger.

Typically, the evolution of the CBC will undergo three phases. The first phase is the *inspiral*, and is roughly the same for all the objects: the two bodies are in quasi-circular orbit and lose energy through GW emission. In first approximation, it lasts until the system reaches the ISCO. Then, there is the merger, where the two objects collide and form the final object, followed by the post-merger regime, which depends on the final object's nature. For BBH mergers, the remnant body is a Kerr BH. It is left in an excited state and undergoes ring-down, where the BH loses energy through GW emission. For an NSBH, the final object is also a Kerr BH, but the expected ring-down is well outside of the LVK detection band [66]. For BNSs the remaining object is either a massive NS or a BH. If it is a BH, then the ring-down occurs, whilst a resultant NS will also lose energy by other processes dependent upon the NS EoS.

The evolution of binaries is described using *waveform approximants*¹. Computing all the details of the process requires solving the EFEs (2.8) entirely through numerical means. This is computationally very expensive. Therefore, one uses other simplifications to describe the GW emission from the binary with a given accuracy, leading to different waveform approximant families.

Before plunging into the details, let us note that a spinning BBH is described typically by 15 parameters: two mass parameters, 6 spin parameters, the luminosity distance, the phase of coalescence, the merger time, the inclination, 2 angles giving the sky location (typically the right ascension and declination), and the polarization angle. For each NS present in the binary (at least) one extra parameter is needed; tidal deformability. It

¹While it is well understood that these are approximations, we loosely call them waveforms or waveform approximants without distinction.

describes how the matter composing the NS is affected by the gravitational fields in the system. Because the relation between the tidal deformability and the mass is determined by the EoS, measuring them accurately for such systems would help constrain the EoS and unravel how matter behaves under these extreme conditions. So, for precessing NSBHs, one has 16 parameters, and 17 parameters are required for BNSs.

This chapter explains the principal waveform families used to analyze GWs, outlining the approximations used to describe the GW emission in the different cases.

3.1 Modeling the Inspiral: the Post-Newtonian Formalism

The inspiral part of the signal can be represented more accurately by using the *post-newtonian* (PN) formalism [67–69]. The main idea is to develop the phase and time (or related quantities, such as their derivatives) in terms of increasing power of v/c and compute the coefficients for each power.

For the time domain, one starts from the energy \mathcal{E} and the luminosity \mathcal{L} (see Ref. [70] for their expression) expanded in terms of $x = (GM\omega/c^3)$, the characteristic frequency of the binary. The energy is known up to 3.5 PN order, and the luminosity up to the third PN order, where the n^{th} PN order corresponds to a term in x^{2n} . From these relations, one can describe the phase and time evolution of the binary using different truncated expansions and equations, leading to a sum of coefficients describing the phase and time evolution as a function of x . In essence, each of the methods develops the energy and the luminosity – separately or together by taking the ratio $\mathcal{F}/(d\mathcal{L}/dx)$ –, and truncates the expression at the desired order. One then transforms the equations to obtain a system of equations describing the binary evolution as a sum of coefficients with increasing PN contributions. For example, one can directly have [70]

$$\phi_{\text{PN}}^{\text{Approx}}(x) = \phi_{\text{ref}}^{\text{Approx}} + \phi_{\text{N}}^{\text{Approx}} \sum_{k=0}^{\text{PN}} \hat{\phi}_k(x) x^k \quad (3.1)$$

$$t_{\text{PN}}^{\text{Approx}} = t_{\text{ref}}^{\text{Approx}} + t_{\text{N}}^{\text{Approx}} \sum_{k=0}^{\text{PN}} \hat{t}_k(x) x^k, \quad (3.2)$$

where $\phi_{\text{ref}}^{\text{Approx}}$ and $t_{\text{ref}}^{\text{Approx}}$ are reference quantities, $\phi_{\text{N}}^{\text{Approx}}$ and $t_{\text{N}}^{\text{Approx}}$ are prefactors to the sum, and $\hat{\phi}_k(x)$ and $\hat{t}_k(x)$ are the coefficients one seeks to compute. All these quantities depend on the waveform approximant, and the truncation methods used. The expressions given in Eqs. (3.1) and (3.2) correspond to those used for the *TaylorT2* waveform approximant, and the coefficients are known up to 3.5 PN order [70]. This waveform is part of the larger *Taylor* waveform family where each waveform approximant

is denoted *TaylorXN*², where X can be T or F depending on whether the waveform is computed in the time or the frequency domain, and N is a number differentiating the different ways to solve the equations determining the phase and time evolution [70]. We refer the reader to Ref. [70] for a more in-depth explanation of the *Taylor* waveforms, with other approximants found with different resummation methods.

The method outlined above is applied in the time domain. However, GW data analyses are generally performed in the frequency domain. Doing the computation in the time domain and then performing a Fourier transform to get to the frequency domain is inefficient. Therefore, it is preferable to directly perform the PN expansion in the frequency domain. Doing this by transforming the expressions for *TaylorT2* into usable ones in the frequency domain leads to the *TaylorF2* waveform approximant. It is obtained through the *stationary phase approximation* (SPA) [72]. In this approximation, one employs the fact that when integrating a function of the form $Ae^{-ig(t)x}$ over t , the main contribution to the integral comes from the part where the phase is stationary, hence $\dot{g} = 0$. When going from a time-domain waveform $h(t)$ to a frequency-domain waveform $\tilde{h}(f)$ using the SPA, the mean contribution for a given f corresponds to the *saddle point* time $t_s(f)$ defined by the condition $\dot{\Phi}(t_s(f)) - 2\pi f = 0$. Under this approximation, the waveform takes the generic form [70]

$$\tilde{h}(f) \simeq \sqrt{\frac{\pi}{2}} \frac{A(t_s(f))}{\sqrt{\ddot{\Phi}(t_s(f))}} e^{i\Psi(f)}, \quad (3.3)$$

where $\Psi(f) = 2\pi f t_s(f) - \Phi(t_s(f)) - \pi/4$. Both A and Ψ can be expressed as function of \mathcal{E} and \mathcal{L} (or their derivatives). As for *TaylorT2*, one expends $\mathcal{L}/(\mathrm{d}\mathcal{E}/\mathrm{d}x)$ and truncates it at the 3.5 PN order. This gives an expression for the waveform where $A \propto g(\alpha, \delta, \iota, \psi) \mathcal{M}_c^{5/6} f^{-7/6}$, where (α, δ) are the right ascension and declination, ι is the inclination of the binary's plane compared to the line of sight, ψ is the polarization angle, and \mathcal{M}_c is the chirp mass (2.50), and the phase term $\Psi(f)$ is known up to 3.5 PN order (see Ref. [70] for the full coefficients). Knowing these terms, Eq. (3.3) can be solved to have a frequency-domain representation of the waveform.

Additionally, one can also want to include spin effects in the waveform (neglected in the explanations and approximants introduced above). Spins are dynamic variables, and they will undergo precession, leading to the precession of the angular momentum to compensate for that of the spins. In turn, this leads to a modulation of the inclination angle over time, and one sees amplitude modulations in the signal's polarization. These effects lead to additional terms in the various expressions accounting for spin-orbit interactions, the coupling between the spins, and the self-interaction for individual spins [71]. For example, such effects are accounted for in the *SpinTaylorT4* [70] and *SpinTaylorF2* [73] waveforms, in the time and the frequency domain, respectively. A comparison between a non-precessing and a precessing GW waveform is shown in Fig. 3.1.

²SpinTaylorXN is also possible. These are waveforms in which the spin effects are included, see Ref. [71] for more details on their inclusion.

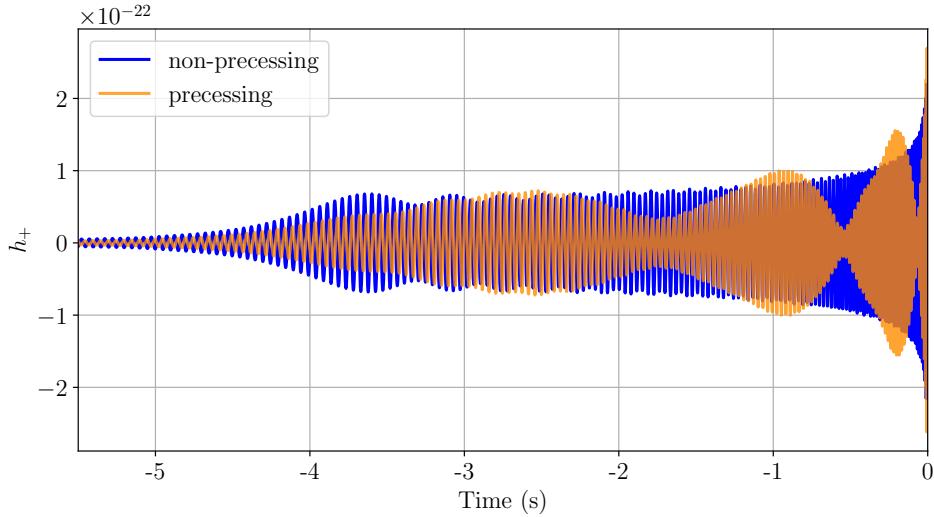


Figure 3.1: Representation of the h_+ polarization for a BBH with and without spin. The inclusion of spin effects leads to a beating pattern in the signal due to the precession of the spin angles leading to changes in the inclination angle, modulating the apparent amplitude.

3.2 Modelling the Full Signal: Numerical Relativity

The waveform introduced in the previous section ended before the merger, around the ISCO frequency. However, the merger and the post-merger regions of the waveforms also contain crucial information. While some methods are developed to find approximate solutions for these regimes, only *numerical relativity* (NR) can provide complete solutions to the EFEs since no finite-order truncation is required. Unfortunately, this process is computationally expensive, needing up to months on supercomputers to generate a few cycles in the inspiral and merger regimes.

Nevertheless, NR has been a growing field in the last 20 years, with the first successful modeling of a non-spinning BBH with equal component masses in 2005 [74]. Since then, many codes have been developed and are capable of generating waveforms for a large variety of CBCs (examples can be found in Refs. [75–81]), and catalogs of numerical NR waveforms are publicly available (for examples see Refs. [82–88]). Even if they are not practicably usable for data analysis, such waveforms are valuable. They are used as models to build, calibrate and assess the error made when using other types of waveforms. Additionally, NR can model the entire environment of a BNS (or any CBC merger), making it a powerful tool for studying the link between EM and GW emissions in such systems (see Ref. [89] for an example).

A specific category of waveforms, called *surrogates* [90–93], can directly follow from NR. Their main idea is to take the waveforms obtained through NR and build an

interpolator between them to cover the possible parameter space and get waveforms faster for a given set of parameters, making their use for data analysis possible. For example, *NRSur7dq2* [94] interpolates NR waveforms and describes precessing BBHs. Its precision is relatively close to NR simulations, but it is about $8\times$ faster.

3.3 Effective One-Body Waveforms

The $(S)EOB(NR)$ family³ is a waveform family capable of modeling the signal from inspiral to ring-down. The main ingredient for these waveforms is the *effective one-body* (EOB) formalism. In the Newtonian realm, one can recast the problem of two rotating bodies into one body moving in an effective potential. In GR, the transformation is a bit more complicated to do, but the EOB formalism applies the same principle [95–97] by mapping the CBC evolution to the motion of a body in an effective metric

$$g_{\mu\nu} = \text{diag}\left(-A(r), \frac{D(r)}{A(r)}, r^2, r^2 \sin^2(\theta)\right), \quad (3.4)$$

where we work in polar coordinates (t, r, θ, ϕ) , and $A(r)$, $D(r)$ are potential which are Taylor approximated.

Moving to the Hamiltonian space in the effective metric, one can define an effective Hamiltonian depending on the conjugate variables of the position and the phase. This can be solved by doing a PN expansion of the potentials present in the Hamiltonian. The related PN-coefficients are known up to 3.5 PN order [98], and one can also add a *pseudo-4 PN* term which is a fiducial parameter included to have a better agreement with NR waveforms [98]. Using this formalism, one can model the entire waveform from its inspiral to the merger by solving the Hamiltonian equations. The ring-down is then described with a ringing Kerr black hole, and smoothly connected to the rest of the model. The full waveform can describe the evolution from inspiral to ring down. Additionally, the effective potential can be modified to add spin effects (e.g. [99, 100]) or matter effects when NSs are present (e.g [101]).

3.4 Phenomenological Waveforms

The EOB waveforms presented in the previous section have the main disadvantage of being slow to evaluate. This is an issue when one performs parameter estimation, where up to

³The S is added if spin effects are accounted for, and NR is added when the waveforms are calibrated against NR.

millions of template waveforms for the different parameter sets are required. One solution to this is the introduction of reduced-order models in the frequency domain [90, 99, 102]. Another alternative is to use *phenomenological waveforms*, designed to be faster than EOB waveforms while being at least as accurate. The philosophy for these waveforms is to model with simple fit functions the frequency-domain behavior of GWs. The parameters are then tuned to match the EOB and NR waveforms.

For example, we can focus on the *IMRPhenomD* waveform [103, 104] valid for an aligned-spin BBH merger and from which other phenomenological waveforms can be derived. For its dominant effect, the GW waveform depends on two mass parameters, the total mass $M = m_1 + m_2$ and the asymmetric mass ratio $\eta = m_1 m_2 / (m_1 + m_2)^2$. In addition, one needs to describe the spin amplitude evolution, which is done through an effective spin, depending on the individual component spins and η [103, 104].

It can be shown that the GW can be decomposed as a function of spin-weighted spherical harmonics [103–105]. Then, the dominant (2, 2) mode can take the generic form

$$\tilde{h}_{22}(f; \Theta) = A(f; \Theta) e^{-i\Phi(f; \Theta)}, \quad (3.5)$$

where Θ represent the binary parameters. So, we seek to describe the phase Φ and the amplitude A using fitted ansatzes. Therefore, the waveform is split into three regions: the inspiral, where the signal is well represented with a hybridized EOB model⁴, an intermediate region linking the inspiral and the final region – the merger and ring-down –. The two last parts are modeled using NR waveforms. We construct a fitting function for the phase and amplitude evolution for each region.

For the **inspiral**, the model is well approximated with the PN formalism. Therefore, it is modeled based on the *TaylorF2* waveform described in Sec. 3.1, complemented with additional terms to correct for deviations from the EOB model (and the NR models for the late inspiral). The description of the phase is

$$\Phi_{\text{ins}} = \Phi_{\text{F2}} + \frac{1}{\eta} \left(\sigma_0 + \sigma_1 f + \sigma_2 f^{\frac{4}{3}} + \frac{3}{5} \sigma_3 f^{\frac{5}{3}} + \frac{1}{2} \sigma_4 f^2 \right), \quad (3.6)$$

where the σ_j are fitting parameters and Φ_{F2} is the phase for *TaylorF2* [70].

For the amplitude in the inspiral and the other regions, one factors out the amplitude of the dominant PN contribution $\propto f^{-7/6}$. This leads to the normalization constant

$$A_0 = \sqrt{\frac{2\eta}{3\pi^{\frac{1}{3}}}} f^{-\frac{7}{6}}. \quad (3.7)$$

Using this, the inspiral amplitude is

$$A_{\text{ins}} = A_{\text{PN}} + A_0 \sum_{j=1}^3 \rho_j f^{\frac{6+j}{3}}, \quad (3.8)$$

⁴This means that the EOB model has been complemented with NR waveforms in the late inspiral.

where A_{PN} is the amplitude in the PN expansion, and ρ_j are the fitting parameters to account for effects not included in the PN formalism.

For the two remaining regions, rather than fitting directly the phase, one constructs the ansatz for its derivative and then integrates the model. This is a way to remove ambiguity on the reference phase and changes the time shift to a simple constant [103].

For the **merger and ring-down**, the characteristic feature in the phase's derivative is a dip. Hence, the value decreases with frequency before increasing again afterward. Therefore, in addition to a polynomial expansion, one gets a damping term which will depend on the *ring-down frequency* of the final object f_{RD} and its *damping frequency* f_{damp} , leading to the phenomenological function

$$\eta \frac{\partial \Phi_{\text{MRD}}}{\partial f} = \alpha_1 + \alpha_2 f^{-2} + \alpha_3 \frac{\alpha_4 f_{\text{damp}}}{f_{\text{damp}}^2 + (f - \alpha_5 f_{\text{RD}})^2}. \quad (3.9)$$

Integrating this leads to the phase for the last region of the waveform

$$\eta \Phi_{\text{MRD}} = \alpha_0 + \alpha_1 f^{-1} + \frac{4}{3} \alpha_3 f^{\frac{3}{4}} + \alpha_4 + \arctan \left(\frac{f - \alpha_5 f_{\text{RD}}}{f_{\text{damp}}} \right), \quad (3.10)$$

where α_0 is a constant of integration. This parameter and α_1 (representing a time-shift in Eq. (3.9)) are determined when imposing smooth transitions with the other regions.

The amplitude is represented by a mixture of a Lorentzian and a decreasing exponential:

$$A_{\text{MRD}} = A_0 \gamma_1 \frac{\gamma_3 f_{\text{damp}}}{(f - f_{\text{RD}})^2 + (\gamma_3 f_{\text{damp}})^2} e^{-\frac{\gamma_2 (f - f_{\text{RD}})}{\gamma_3 f_{\text{damp}}}}. \quad (3.11)$$

If one only considers the Lorentzian component in Eq. (3.11), the peak amplitude happens for f_{RD} , which is not matching the observations in NR waveforms. Using Eq. (3.11), one finds the peak frequency to be

$$f_{\text{peak}} = \left| f_{\text{RD}} + \frac{f_{\text{damp}} \gamma_3}{\gamma_2} \left(\sqrt{1 - \gamma_2^2} - 1 \right) \right|. \quad (3.12)$$

The presence of the ring-down and damping frequencies in Eqs. (3.10), (3.11), and (3.12) is not so surprising since we expect the final object to having repercussions on the behavior of the frequency evolution in the latest stages of the merger.

The **transition region**, linking the two regions described above, has a dominant phase derivative evolution going as f^{-1} . It is corrected by a term in f^{-4} accounting for the observed deviations from the dominant behavior. So,

$$\eta \frac{\partial \Phi_{\text{int}}}{\partial f} = \beta_1 + \beta_2 f^{-1} + \beta_3 f^{-4}. \quad (3.13)$$

Integrating this yields the expression for the phase in the intermediate region

$$\eta\Phi_{\text{int}} = \beta_0 + \beta_1 f + \beta_2 \ln(f) - \frac{\beta_3}{3} f^{-3}, \quad (3.14)$$

where β_0 is a constant of integration and β_1 is a time-shift term. Their values are fixed through the smooth matching condition between the different regions.

Since smoothly linking the first and last regions' amplitude is not valid in all generality, one represents the amplitude in this region as a fourth-order polynomial where the coefficients are fixed by requiring matching conditions with the inspiral and merger regions:

$$A_{\text{int}} = A_0(\delta_0 + \delta_1 f + \delta_2 f^2 + \delta_3 f^3 + \delta_4 f^4). \quad (3.15)$$

Therefore, the δ_j coefficients are solutions to the system of equations

$$A_{\text{int}}(f_1) = A_{\text{ins}}(f_1) \quad (3.16)$$

$$A_{\text{int}}(f_2) = A_{\text{NR}}(f_2) \quad (3.17)$$

$$A_{\text{int}}(f_{\text{peak}}) = A_{\text{MRD}}(f_{\text{peak}}) \quad (3.18)$$

$$\frac{dA_{\text{int}}}{df}(f_1) = \frac{dA_{\text{ins}}}{df}(f_1) \quad (3.19)$$

$$\frac{dA_{\text{int}}}{df}(f_{\text{peak}}) = \frac{dA_{\text{MRD}}}{df}(f_{\text{peak}}), \quad (3.20)$$

where $Mf_1 = 0.014$, f_{peak} is defined in Eq. (3.12), and $f_2 = 0.5(f_1 + f_{\text{peak}})$, and $A_{\text{NR}}(f_2)$ is the amplitude of the NR waveform at f_2 .

Accounting for the parameters fixed through matching conditions, we have 17 phenomenological parameters needing to be mapped to the physical ones. This can be done using a polynomial function of the asymmetric mass and effective spin – see [104] for the details.

Once the different regions are obtained, they are “stitched” together using two different types of window functions:

$$\Xi_{f_0}(f) \equiv \Xi(f - f_0) = \begin{cases} -1 & \text{if } f < f_0 \\ 1 & \text{if } f \geq f_0 \end{cases} \quad (3.21)$$

and

$$\Xi_{f_0}^{\pm} = \frac{1}{2}(1 \pm \Xi(f - f_0)). \quad (3.22)$$

The final waveform is then

$$\tilde{h}_{22}(f) = A_{\text{IMRPD}}(f) e^{-i\Phi_{\text{IMRPD}}(f)}, \quad (3.23)$$

with

$$\Phi_{\text{IMRPD}}(f) = \Phi_{\text{ins}}(f)\Xi_{f_1}^-(f) + \Xi_{f_1}^+(f)\Phi_{\text{int}}(f)\Xi_{f_2}^- + \Xi_{f_2}^+(f)\Phi_{\text{MRD}}(f), \quad (3.24)$$

where the phase parameters for the different regions are given by Eqs. (3.6), (3.14) and (3.11), $Mf_1 = 0.018$, and $f_2 = (1/2)f_{\text{RD}}$, and

$$A_{\text{IMRPD}}(f) = A_{\text{ins}}\Xi_{f_1}^-(f) + \Xi_{f_1}^+(f)A_{\text{int}}(f)\Xi_{f_2}^- + \Xi_{f_2}^+(f)A_{\text{MRD}}(f), \quad (3.25)$$

where the amplitude terms are given in Eqs. (3.8), (3.15) and (3.11), and the frequencies are such that $Mf_1 = 0.014$ and $f_2 = f_{\text{peak}}$.

With this, *IMRPhenomD* [103, 104] is a frequency-domain waveform for non-precessing BBHs that is rapid to compute. Conversely, the waveform is calibrated to be accurate, with accuracies similar to the EOB waveforms in its inspiral, and close to NR in the other regions. However, it seems to lack some physics due to its aligned spin assumption. Nevertheless, it is an important step towards fast and accurate waveform models as one can transform non-precessing waveforms into precessing ones by “twisting up” the first. In essence, this is done by applying a set of time-dependent rotation angles $(\beta(t), \gamma(t), \varepsilon(t))$ such that [106, 107]

$$h_{2m}^{\text{precession}} = e^{-im\varepsilon(t)} \sum_{m'=-2}^2 e^{im'\gamma(t)} d_{mm'}^2(\beta(t)) h_{mm'}(t), \quad (3.26)$$

where $h_{mm'}(t)$ are the non-precessing dominant mode (described by *IMRPhenomD*, for example) $d_{mm'}^2$ are Wigner d-matrices [108, 109]. Different methods for computing the angles exist. Applying the one from Ref. [107] converts *IMRPhenomD* to *IMRPhenomPv2*.

For BNS waveforms, one also needs to account for the tidal deformability of the objects as the NSs get deformed through the gravitational interaction, leaving an imprint in the GW signal. It is possible to add these matter effects into the phenomenological waveforms by assuming a decoupling between the phasing effect due to the tidal deformability and the other effects determining the CBC’s phase evolution [110]. Following the principles defined above, one can define a fitting formula for the phase effect related to tidal deformability and fit it to numerical BNS waveforms. This leads to the *IMRPhenomD_NRTidal* waveform [110], where the precession and self-spin effects are neglected, and the *IMRPhenomPv2_NRTidal* waveform [111], where the spin dynamics and the tidal deformability are included.

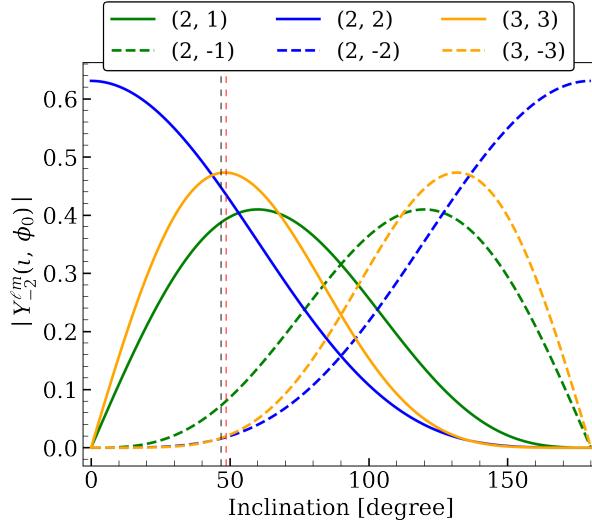


Figure 3.2: Representation of the absolute value of the spin-weighted spherical harmonic as a function of the inclination angles. The black and red dashed lines represent the inclinations for GW190412 [41] and GW190814 [44], respectively. This explains why these events have visible HOMs. Figure adapted from [114].

3.5 Higher-Order Modes in Gravitational Waves

Up to now, our discussion has focused on the dominant mode of the spherical harmonic decomposition. However, when the systems are made of dissimilar masses and/or have a significant inclination, other modes are important. Fig. 3.2 gives a representation of the strength of some sub-dominant modes depending on the binary’s inclination. Also, recall from Chapter 1 that there are now already two binaries with confirmed detection of such higher-harmonics: GW190412 [41] and GW190814 [44]. In addition, observing such modes can lead to additional tests of GR [112–114] and a better identification of strongly-lensed GWs [115–117].

The three approaches detailed in the previous sections can deal with HOMs. As was the case above, the phenomenological models rely on the EOB models, which themselves depend on the PN expansions. HOMs effects can also be accounted for in addition to precession and matter effects explained in Sec. 3.4.

First, the waveform is expressed in terms of spin-weighted spherical harmonics [105]

$$h(f; \Theta, \theta, \phi) = \sum_{l=2}^{+\infty} \sum_{m=-l}^l h_{lm}(f; \Theta)^{-2} Y_{l,m}(\theta, \phi), \quad (3.27)$$

where Θ represents the binary’s intrinsic parameters, and (θ, ϕ) represents the orientation

with respect to the observer. The different modes of the GW are

$$h_{lm} = \frac{2M\eta}{D_L} x^2 \sqrt{\frac{16\pi}{5}} \hat{h}_{lm} e^{im\Psi}, \quad (3.28)$$

where the values of \hat{h}_{lm} for a precessing BBH can be found in Refs. [105, 118].

Using the SPA introduced in Sec. 3.1 adapted for a generic precessing BBH [105], one can re-write the h_{lm} from Eq. (3.28) in the Fourier domain as [118]

$$\tilde{h}_{lm}(f) = \frac{\pi M^2}{D_L} \sqrt{\frac{2\eta}{3}} V_m^{\frac{-7}{2}} e^{-i(m\Psi_{\text{SPA}}(V_m) + \pi/4)} \hat{H}_{lm}(V_m), \quad (3.29)$$

where Ψ_{SPA} is the phase found under the SPA approximation (and can also account for spin effects, see Refs. [105, 118] for example), the $\hat{H}_{lm}(V_m)$ have been computed up to $l = 6$ [118], and $V_m = (2\pi M f_k/k)^{1/3}$, with $f_k = k\omega_{\text{orb}}/(2\pi)$.

One can then use the same resummation methods as those described in Sec. 3.3 for the PN expressions, leading to an EOB description accounting for precession and HOMs [119]. An example EOB waveform accounting for those effects is the *SEOBNRv4PHM* waveform [119].

Once the EOB waveform is designed, one can also make a phenomenological waveform for non-precessing (*IMRPhenomHM* [120]) or precessing (*IMRPhenomXPHM* [121]) BBH with HOMs. The procedure followed is the same as in Sec. 3.4, where the inspiral part is calibrated on the hybridized EOB waveforms and the rest using NR waveforms (now also upgraded to contain the full HOM content).

4

GRAVITATIONAL-WAVE DATA AND ITS ANALYSIS

In Chapter 1, we have briefly explained some of the challenges encountered in developing detectors for GW detection. Due to the small amplitude of the signal, the interferometers require very high sensitivity. Therefore, they are subject to many noise sources. In this chapter, we detail how one can extract a faint signal and its characteristics from such large noise backgrounds.

4.1 The Noise and its Challenges

During the discussion about the development of the detectors and the significant issues faced, we mentioned that one of the main bottlenecks for getting the required sensitivity for detection is the noise, such as the seismic noise and the shot noise on the mirrors [24], for example. Many studies and techniques are still ongoing to better understand and eliminate the noise sources [122]. The noise reduction was key to the first GW detection, and its suppression is crucial to improve GW astronomy [31]. However, noise is a part of the equation, and one has to deal with it in the analysis framework.

When a GW enters a detector's sensitive band, the recorded data $d(t)$ is

$$d(t) = n(t) + h(t), \quad (4.1)$$

where $n(t)$ represents the noise component, and $h(t)$ is the GW signal, often described with waveforms as presented in Chapter 3. From this chapter, we have seen we have relatively good models for GWs. However, we also need to model the noise if we want to extract the signal.

Typically, one represents the noise as stationary and Gaussian. Therefore, its characteristics do not change over extended periods, and it has an average value of zero. In the frequency domain, these assumptions entail

$$\langle \tilde{n}^*(f) \tilde{n}(f') \rangle_{\text{avg}} = \frac{S_n(f)}{2} \delta(f - f'), \quad (4.2)$$

where the tilde means the Fourier transform, $\langle \cdot \rangle_{\text{avg}}$ represents the ensemble average, and

$S_n(f)$ is called the power spectral density (PSD) and represents the noise content of the detector around a given frequency. In practice, the noise characteristics are computed over a finite duration T , and the average becomes a time average:

$$\langle |\tilde{n}(f)|^2 \rangle_T = \frac{S_n(f)}{2T}. \quad (4.3)$$

In the Fourier space, a frequency bin $\Delta f = 1/T$. Therefore, Eq. (4.3) shows $S_n(f)/2$ is the variance of the noise in a given frequency bin.

Fig. 4.1, left panel, represents the average PSD of the LIGO and Virgo detectors during the O3 observation run. The peaks in the curve are typically due to noise artifacts in the instrumentation and can limit the sensitivity. As a rule of thumb, when the strain of the signal is above these curves, it can be detected, and we can extract information from it. Assuming the noise is Gaussian and stationary, we can design a filter to find signals in the data. In reality, the noise can also have short-lived variations, called *glitches*, where the noise curve is modified for a variable period of time [123–126]. They need to be mitigated to avoid biases in the events’ characterization [126]. For example, GW170817 suffered from a glitch during its inspiral in the LIGO-Livingston detector (see the right panel of Fig. 4.1). The part of the data containing the glitch was removed before analyzing the data [37]. Finally, we can also mention another source of uncertainty: the possible presence of missed or undetectable GW signals.

Even if we know the noise is not entirely stationary and Gaussian, searches and analysis methods are designed for such a noise. This leads to sub-optimality and slight biases in the recovered information. Still, the noise behaves well enough to permit the analyses under these assumptions.

4.1.1 Extracting a Gravitational-Wave Signal from Noise

In Eq. (4.1), $|h(t)| \ll |n(t)|$, requiring an adapted method to extract the signal from the noise. This method is called *matched filtering*, and relies on the *Wiener-Khinchin theorem*, linking the auto-correlation function of a stationary process to its power spectrum [127].

If $K(t)$ is an unspecified filter applied to the data, we can define

$$\hat{d} = \int_{-\infty}^{+\infty} dt d(t) K(t). \quad (4.4)$$

We can use Eq. (4.4) to define a quantity comparing the noise-only and the noise + GW cases. It is called the *signal-to-noise ratio* (SNR, written ρ) and is the ratio between \hat{d} ’s expected value when a GW is present in the data (S) and \hat{d} ’s root-mean-square value when there is only noise (N).

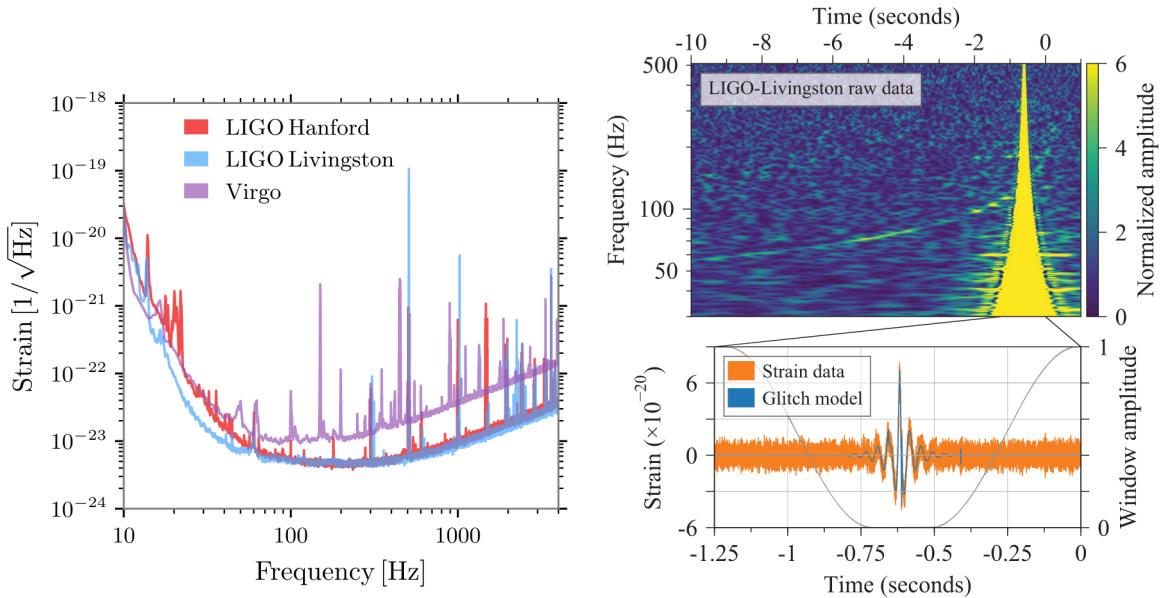


Figure 4.1: *Left:* The mean PSD for the LIGO and Virgo detectors during the third observation run [40]. The peaks are due to instrumental noises. *Right:* The time-frequency map for the GW170817 signal in LIGO-Livingston with the glitch present in the inspiral, and the glitch's strain [37]. The part of the data containing the glitch had to be removed before analyzing the data.

When a signal is present in the data

$$\begin{aligned} S = \hat{d} \Big|_{d(t)=n(t)+h(t)} &= \int_{-\infty}^{+\infty} dt \langle n(t) + h(t) \rangle_{\text{avg}} K(t) \\ &= \int_{-\infty}^{+\infty} dt h(t) K(t) = \int_{-\infty}^{+\infty} df \tilde{h}(f) \tilde{K}^*(f), \end{aligned} \quad (4.5)$$

where one should recall $\langle n(t) \rangle_{\text{avg}} = 0$ by the noise properties.

On the other hand, when only noise is present in the data

$$\begin{aligned} N^2 &= \langle \hat{d}^2 \rangle - \langle \hat{d} \rangle^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} dt dt' K(t) K(t') \langle n(t) n(t') \rangle \\ &= \int_{-\infty}^{+\infty} df \frac{S_n(f)}{2} |K(f)|^2, \end{aligned} \quad (4.6)$$

where the square root of this expression is the denominator of the SNR.

Defining the *noise weighted inner product*

$$\langle p | q \rangle = \mathbb{R} \left(2 \int_{-\infty}^{+\infty} df \frac{\tilde{p}(f) \tilde{q}^*(f)}{S_n(f)} \right), \quad (4.7)$$

the SNR is

$$\rho = \frac{\langle k|h \rangle}{\sqrt{\langle k|k \rangle}}, \quad (4.8)$$

where

$$\tilde{k}(f) = \frac{K(f)S_n(f)}{2}. \quad (4.9)$$

By definition, the presence of a signal in the data is the most obvious when the SNR is maximized. From Eqs. (4.8) and (4.9), we see that the maximum value is reached when

$$\tilde{k}(f) \propto \frac{\tilde{h}(f)}{S_n(f)}. \quad (4.10)$$

The filter with this property is called the *Wiener filter*, and it leads to the optimal SNR

$$\rho_{\text{opt}} = \sqrt{\langle h|h \rangle}. \quad (4.11)$$

It is the highest value of the SNR one can get, and is obtained when computing the inner product of the signal with itself.

The SNR is a way to find a signal in the data. For example, we can say that for a given noise, the average SNR value is X and that we are confident of the presence of a GW in the signal if it leads to an SNR $Y \gg X$.

For GW data analysis, the GW observed for each detector is dependent on the detector location and the sky location through the beam-pattern functions (2.26). For a given GW in the sky frame, the observed wave in the different detectors is different. So, we can compute the SNR for each of them by taking the inner product of the GW in the specific detector frame with the data for the said interferometer. The significance of the GW signal in the detector network is then expressed using the *network SNR* [128]:

$$\rho_{\text{net}} = \sqrt{\sum_{j=1}^{N_{\text{det}}} \rho_j^2}, \quad (4.12)$$

where N_{det} is the number of detectors, and ρ_j is the SNR for detector j .

4.1.2 Wiener Filters for Gravitational-Wave Searches

In Eqs. (4.8) and (4.10), the optimal filter is dependent on the signal itself, which can appear problematic. A priori, when searching for a GW, one does not know the signal in advance. To circumvent this, matched filtering is applied using a *template bank*: a collection of model waveforms placed to have a minimal difference between them such that we should not miss too many signals [129].

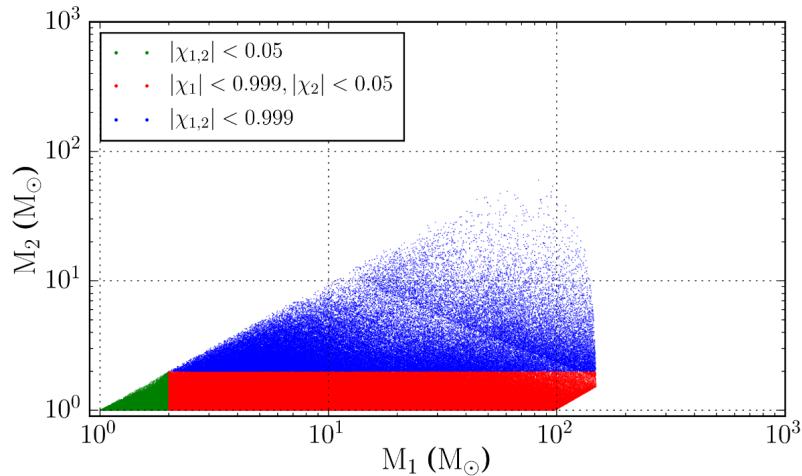


Figure 4.2: Template bank used during the O2 observation run [129] for the GstLAL pipeline [130, 131] represented in the component mass plane. Blue is the BBH systems, red is the NSBH systems, and green is the BNS ones. Smaller masses are populated more densely, showing that a given absolute change in intrinsic parameters makes for a bigger loss in SNR when the masses are small.

Typically, one generates BBH, BNS, and NSBH templates using some waveform. The intrinsic parameter space is covered such that we “do not lose too much signal”. To quantify the similarity between waveforms, one defines the *match*

$$\mathcal{M}(h_1, h_2) = \max_{\phi} \left(\frac{\langle h_1 | h_2 \rangle}{\sqrt{\langle h_1 | h_1 \rangle \langle h_2 | h_2 \rangle}} \right), \quad (4.13)$$

where $\langle . | . \rangle$ is the weighted inner product (4.7), and h_1, h_2 are two waveforms we want to compare. If $h_1 = h_2$, the match is one, and it decreases as the waveforms become more dissimilar. The match translates the percentage in SNR kept from one waveform to the other. The templates are placed in the bank such that the match between one template and its neighbors is not lower than a given value, for example, 0.97 in [129] used for the second observation run¹. Fig. 4.2 gives a representation of the template bank used for O2 [129] in the GstLAL GW search pipeline [130, 131]. An interesting observation is that the bank is populated more densely for lower mass systems. One can understand this as a manifestation of a bigger difference between two signals for a given absolute parameter variation, which occurs more easily for longer waveforms, *i.e.* at lower masses. This also impacts the parameter estimation process since it represents our ability to distinguish between two neighboring events.

Once the bank is made, one can take time windows in the data and apply matched

¹To be precise, this is the coverage required for online searches, hence performed during data taking. For offline searches, done after the observation run is completed and one has more time to do the analysis, the template bank is enhanced with more waveforms [129].

filtering for each waveform in the pool and see if some present a peak in SNR. However, GW detector noise is not Gaussian nor stationary, and extra steps are needed to ensure the signal’s origin. These and the waveforms used in the bank are different from one pipeline to the other. In addition to GSTLAL [130, 131], there is also PyCBC [132–137], and MBTA [138, 139]².

As an example, we can see how GSTLAL identifies a candidate [40, 130, 131]. First, it does match filtering for each template in each detector. A candidate event is found if it has an SNR higher than four in each online detector for the same template. Then, a background is made by sliding the data from one detector relative to the other and evaluating the SNR. This breaks the coincidence between the detectors, ensuring that the evaluated SNR and triggers are not due to a genuine GW event. A likelihood is then built based on the network matched-filter SNR, the background-informed vetoes, the measured phase and time differences between the detectors, and information about short non-Gaussian noise artifacts [144]. One then classifies the triggers by decreasing values of likelihood ratio which compares the probabilities of the trigger originating from a GW event and the trigger being due to noise, leading to a false-alarm rate (FAR) representing the time needed for noise-only data to produce a trigger with the same significance. Additionally, one also computes the p_{astro} [145, 146] – introduced in the second half of the third observing run [40] –, giving the probability that the event has an astrophysical origin. If the FAR is low enough, and $p_{\text{astro}} > 0.5$, the event is analyzed further. The same procedure is followed by the other pipelines, except they use different ways to classify the triggers and ensure their origin (see Appendix D in [40] for a brief overview of the different methods).

4.2 Bayesian Statistics, the Key to Data Analysis

The previous section shows how one finds a GW signal in the data. However, it only gives access to the template(s) matching the data, leading to point-approximate values for the GW parameters. Typically, after the signal is identified, it is further analyzed to grasp its characteristics better. This step is called parameter estimation. In addition, when checking whether the GW has particular characteristics (for example, deviations from GR, lensing, …), one needs to compare the result of the analysis under two hypotheses (for example, GR versus non-GR, lensed or unlensed, …). The key to all these concerns is Bayesian statistics.

²Note that there are also search pipelines not based on template banks. For example, SPIIR is equivalent to matched filtering but uses infinite response filters to model the signals [140, 141], and cWB does not use a model and looks for coherent power excess in the different detectors [142, 143].

4.2.1 Bayes' Theorem

Most of the concepts in the Bayesian framework result from two fundamental rules:

- The *product rule*: $P(A, B) = P(A|B)P(B)$, where $P(A, B)$ is the probability to have A and B , $P(A|B)$ is the probability to have A given B , and $P(B)$ is the probability to have B ;
- The *complement rule*: $P(A) + P(\neg A) = 1$, where $P(\neg A)$ is the probability of not having A .

In particular, using the product rule, one can find *Bayes' theorem*:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \quad (4.14)$$

which is fundamental for data analysis. However, in that context, one often works under a given hypothesis, changing the expression to

$$P(A|B, \mathcal{H}) = \frac{P(B|A, \mathcal{H})P(A|\mathcal{H})}{P(B|\mathcal{H})}, \quad (4.15)$$

where $P(B|A, \mathcal{H})$ is the probability of having B given A under the hypothesis \mathcal{H} . Each term in the right-hand side of Eq. (4.15) has a signification. $P(B|A, \mathcal{H})$ is called the *likelihood*, and it is typically a computable function. $P(A|\mathcal{H})$ is the *prior* on A and represents the belief one has on A before running the experiment. Finally, $P(B|\mathcal{H})$ is the *evidence*, and it is a normalization constant obtained by integrating the numerator over all values of A .

It is useful to specify Bayes' theorem for some specific scenarios related to data analysis. First, one often wants to assess the probability of being under a certain hypothesis \mathcal{H} for the observed data d . For example, in GW data analysis, given data are we observing a GW (\mathcal{H}_{GW})? For an arbitrary hypothesis \mathcal{H} one has

$$P(\mathcal{H}|d) = \frac{P(d|\mathcal{H})P(\mathcal{H})}{P(d)}, \quad (4.16)$$

It is usually easier to determine the probability of having the observed data under a certain hypothesis (likelihood) than the probability of having the hypothesis given the data. $P(\mathcal{H})$ represents the prior probability to be under the hypothesis \mathcal{H} .

We see that using Bayes' theorem somewhat resembles a process of updating what one believes. First, one thinks there is a certain chance of being in a given situation (prior on the hypothesis). Then, one runs an experiment to collect data and evaluates the likelihood. Based on the likelihood and the evidence, one updates one's belief to be in the scenario \mathcal{H} accounting for the data.

Often, one analyses the data using a model including parameters. In such a case, one is interested in having the value of the model's parameters $\boldsymbol{\theta}$ given the data and the hypothesis:

$$P(\boldsymbol{\theta}|d, \mathcal{H}) = \frac{P(d|\boldsymbol{\theta}, \mathcal{H})P(\boldsymbol{\theta}|\mathcal{H})}{P(d|\mathcal{H})}, \quad (4.17)$$

which is the expression one evaluates when determining the values of the parameters describing a GW event. $P(\boldsymbol{\theta}|d, \mathcal{H})$ is a distribution for the model parameters $\boldsymbol{\theta}$ when accounting for the data and the hypothesis, and is called the *posterior*.

For GW data analysis, the likelihood – linking the data, the hypothesis, and the model parameters – in the frequency domain is

$$\mathcal{L} = P(d|\boldsymbol{\theta}, \mathcal{H}) \propto \exp \left[-\frac{1}{2} \langle d - h(\boldsymbol{\theta}) | d - h(\boldsymbol{\theta}) \rangle \right], \quad (4.18)$$

where $\langle . | . \rangle$ is the weighted inner product (4.7), and the expression follows from the assumption that the noise n is stationary and Gaussian, and $d = n + h$. For a given waveform model and set of binary parameters, $h(\boldsymbol{\theta})$ can be evaluated, and one can compute the likelihood, making it possible to find the posterior distributions. In Sec. 3.4, we mentioned that phenomenological models are introduced to be faster. Eq. (4.18) shows why it is needed: we need to evaluate the likelihood by generating the waveform and subtracting it from the data. Moreover, it needs to be evaluated numerous times to explore the parameter space, making the waveform generation a major bottleneck when doing parameter estimation.

4.2.2 Summarizing the Posteriors

The posteriors obtained using Eq. (4.17) can be multi-dimensional. This is very hard to visualize, and one is often interested in the probability density for one parameter (or a sub-group of parameters). This is obtained by performing *marginalization*. If $\boldsymbol{\theta} = \{\theta_1, \theta_2, \dots, \theta_N\}$, the posterior for θ_1 is

$$p(\theta_1|d, \mathcal{H}) = \int d\theta_2 \dots d\theta_N p(\boldsymbol{\theta}|d, \mathcal{H}). \quad (4.19)$$

Fig. 4.3 represents an example of 2-D and 1-D posteriors in the form of a corner plot. The top and bottom right panels represent the marginalized posteriors for the right ascension and declination of a GW event. The bottom left panel is the joint posteriors, where the different colors represent different *confidence interval*, which are a way to summarize information present in the data.

For a given posterior, one can compute summary information that translates some

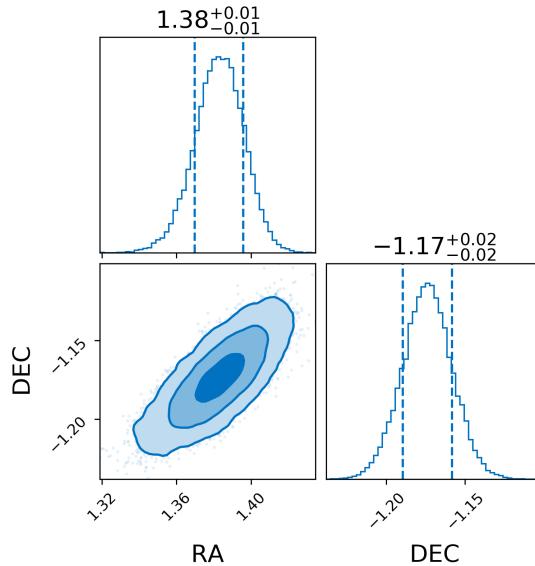


Figure 4.3: Representation of the right ascension (RA) and declination (DEC) of an event as a corner plot. The side panels show the marginalized distribution for the two parameters, while the central one represents the joint posterior distribution. The different color levels represent different confidence regions (0.68, 0.95, and 0.997) in the 2D posterior. In the 1D posteriors, the dashed lines represent the limit of the 0.68 confidence interval. The values written above the 1D posteriors are the median values for the parameters and their values at the edges of the 0.68 confidence interval.

characteristics of the distribution, for example, the *mean value*

$$\bar{\theta} = \int_{\theta_{\min}}^{\theta_{\max}} d\theta p(\theta|d, \mathcal{H})\theta. \quad (4.20)$$

This quantity is not ideal in some cases. For example, if a posterior is bimodal, the mean value can fall in a region where the samples are not present. Therefore, one often prefers to use the *median* value θ^{med} of the distribution, such that

$$\int_{\theta_{\min}}^{\theta^{\text{med}}} p(\theta|d, \mathcal{H}) = 0.5. \quad (4.21)$$

Hence, it is the value for which half the probability density is present. θ^{med} is a value present in the data set. If the distribution is unimodal and symmetric, the mean and the median are close.

In addition to a representative point of the distribution, it is often valuable to know how (un)certain we are of the estimation. Intuitively, a wider posterior means that the parameter is compatible with more values than a tighter one. It is often summarized

using a *confidence interval*

$$\gamma_X = \int_{\boldsymbol{\theta}_{\text{low}}}^{\boldsymbol{\theta}_{\text{high}}} d\boldsymbol{\theta} p(\boldsymbol{\theta}|d, \mathcal{H}), \quad (4.22)$$

where X represents the fraction of probability density included in the $[\boldsymbol{\theta}_{\text{low}}, \boldsymbol{\theta}_{\text{high}}]$ interval. Often, one computes $\gamma_{0.68}$ and $\gamma_{0.95}$, the one- and two-sigma confidence interval. The dashed lines in the 1D posteriors from Fig. 4.3 are the 1-sigma confidence interval. For the 2D posterior, the darkest region is the 1-sigma confidence region, the slightly lighter one is the 2-sigma confidence region, and the lightest region is the 3-sigma confidence region (corresponding to $\gamma_{0.997}$). Another way of interpreting these confidence intervals/regions is that the true value of $\boldsymbol{\theta}$ has a probability of X of being in γ_X .

4.2.3 Hypothesis Testing

Hypothesis testing – i.e. determining whether it is more likely to be in some scenario or another, summarized by two hypotheses \mathcal{H}_1 and \mathcal{H}_2 – is a crucial aspect of data analysis. For example, one compares the case where there is a GW in the data (\mathcal{H}_{GW}) or only noise (\mathcal{H}_{N}). Another possibility, investigated in this thesis, is to compare the lensed hypothesis (\mathcal{H}_{L}) (the observed GW is lensed) and the unlensed hypothesis (\mathcal{H}_{U}) (the observed GW is not lensed). In all generality, one compares two hypotheses \mathcal{H}_1 and \mathcal{H}_2 using the *odds ratio*

$$\mathcal{O}_{\mathcal{H}_2}^{\mathcal{H}_1} = \frac{p(\mathcal{H}_1|d)}{p(\mathcal{H}_2|d)} = \frac{p(d|\mathcal{H}_1)}{p(d|\mathcal{H}_2)} \frac{p(\mathcal{H}_1)}{p(\mathcal{H}_2)} = \mathcal{B}_{\mathcal{H}_2}^{\mathcal{H}_1} \mathcal{P}_{\mathcal{H}_2}^{\mathcal{H}_1}. \quad (4.23)$$

It is the ratio of the probabilities to be under a hypothesis given the data. The likelihood ratio $\mathcal{B}_{\mathcal{H}_2}^{\mathcal{H}_1}$ is called the *Bayes factor* and is driven by the data. $\mathcal{P}_{\mathcal{H}_2}^{\mathcal{H}_1}$ is the prior odds and is driven by beliefs before any experiment. When the Odds ratio is greater than one, \mathcal{H}_1 is favored. If it is smaller than one, then \mathcal{H}_2 is preferred. When relatively close to one, we cannot choose a hypothesis over the other. The prior odds is often a tricky quantity compared to the Bayes factor as it is not computed based on the data. Therefore, in some cases, one sets $\mathcal{P}_{\mathcal{H}_2}^{\mathcal{H}_1} = 1$ and evaluates the odds based on the Bayes factor.

From Eq. (4.23), it can appear that the favored model will always be the one that fits our data the best. Therefore, if we use an arbitrarily large number of parameters, we expect the model to be favored in the odds ratio. This is not accounting for *Occam's razor* (or *principle of parsimony*), saying that if an additional parameter does not improve the theory's accuracy, then it increases the probability that the theory is wrong. This is a consequence of the increase in prior volume induced by the addition of new parameters and is automatically accounted for in the Odds ratio. This principle also plays a role when searching for GW strong lensing, where lens models have more parameters than unlensed ones but are not automatically favored.

4.2.4 Final Note on Bayesian Statistics

All the examples given in this section considered continuous variables and probabilities. In data analysis, we are often confronted with finite samples with discrete parameter spaces. The rules followed by the different quantities remain the same, and the expression defined above stays valid. Typically, to go from the continuous to the discrete expression, it suffices to replace the integrals with sums. So, for example, the confidence interval (4.22) becomes

$$\gamma_X^{\text{disc}} = \sum_{\boldsymbol{\theta}=\boldsymbol{\theta}_{\text{low}}}^{\boldsymbol{\theta}^{\text{high}}} p(\boldsymbol{\theta}|d, \mathcal{H}), \quad (4.24)$$

where $\boldsymbol{\theta}$ is an ensemble of data points forming the discrete probability density. Similarly, the evidence is computed by summing over the products between the likelihood and the prior for the various points.

4.3 Nested Sampling

Data analysis generally seeks to know the preferred hypothesis and the values of the parameters describing the model. Doing this via brute-force methods, like Monte-Carlo integration is generally unsuccessful because the problems are complex and have high dimensionalities. Therefore, other means are required to evaluate the evidence and get the posterior samples. *Nested sampling* is such a method.

4.3.1 Obtaining the Evidence and Posteriors with Nested Sampling

Because the evidence often requires performing integration over a large parameter space when expressed in terms of the parameters, nested sampling [147] proposes a change in perspective. Instead of relying directly on the physical parameters, it uses the *prior mass*

$$X(\lambda) = \int_{p(d|\boldsymbol{\theta}, \mathcal{H}) > \lambda} d\boldsymbol{\theta} p(\boldsymbol{\theta}|\mathcal{H}) \quad (4.25)$$

to evaluate the likelihood. $X(\lambda)$ is a continuous function decreasing with λ as a higher λ value leads to integration over a smaller portion of the prior probability density. It is bounded between $X(0) = 1$ since the prior is normalized and $X(\lambda_{\max}) = 0$ for λ_{\max} corresponding to the point of maximum likelihood.

Writing the likelihood L , we can inverse Eq. (4.25) to get $L(X) = L(X(\lambda))$. In these

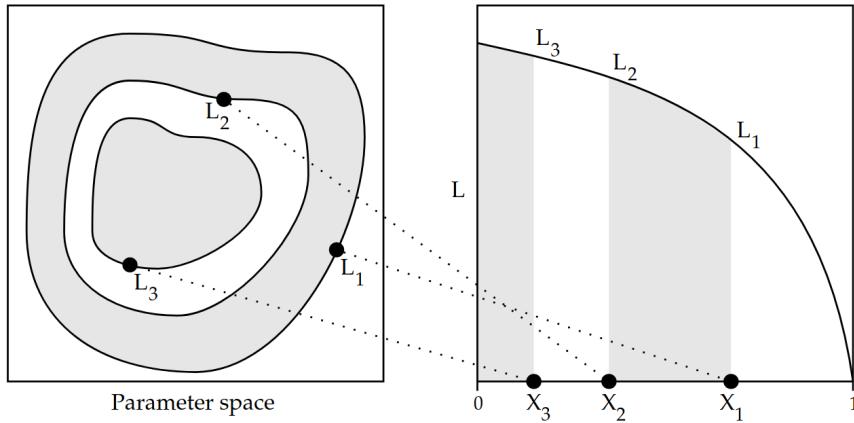


Figure 4.4: Representation of the different steps followed by nested sampling when going from one point to the other [147]. For each iteration, one retains the point of minimum likelihood, draws a new one, and re-orders the samples before starting over. This enables us to explore the parameter space. Figure adapted from [147].

terms,

$$\mathcal{Z} = \int_0^1 dX L(X), \quad (4.26)$$

which has the advantage of being a bounded integral. $L(X)$ is monotonically increasing function, with $L = 0$ for $X = 1$, and $L = L_{\max}$ for $X = X_{\min}$.

To evaluate Eq. (4.26), one needs to have a set of points $\{(L_i, X_i)\}_{i=1,\dots,N}$. Drawing N samples $\{\theta\}_{i=1,\dots,N}$ from the prior $p(\theta|\mathcal{H})$, we can compute the corresponding likelihoods $\{L_i\}_{i=1,\dots,N}$ and priors masses $\{X_i\}_{i=1,\dots,N}$. One can then order the samples by increasing likelihood values (hence by decreasing prior masses)

$$\begin{aligned} 0 < L_1 < L_2 < \dots < L_N, \\ 1 > X_1 > X_2 > \dots > X_N > 0. \end{aligned}$$

This leads directly to a point with a minimum likelihood: (L_1, X_1) , and gives us our first dead point. We then draw a new sample from the prior with the condition that it has a larger likelihood than the previous point. We can then order the samples again and have the next dead point. Thus, we always keep a pool of N samples and each time select the one with the lowest likelihood. This way, we go to increasing likelihood values for each next sample. Fig. 4.4 represents this process, where each point has a smaller likelihood surface than the previous one with a value that is necessarily higher than the one found for the last dead point. So, we explore the likelihood space by going to smaller surfaces with higher likelihoods.

The re-ordering step in the algorithm leads to inefficiency. However, this step can be bypassed by using a probabilistic estimation of the prior mass. Since the set of samples is taken from the prior distribution, their likelihood is such that the sampling is equivalent

to taking the mass prior in a uniform distribution between 0 and 1. So, the probability that the prior mass is drawn with a value below some X' is

$$P(X') = \int_0^{X'} dX = X'. \quad (4.27)$$

If \tilde{X} is the highest mass prior in the samples, we know $\tilde{X} < 1$ and is associated with the smallest likelihood \tilde{L} .

So, one can compute the probability that the surface of the highest mass prior is at a certain value X^* . It is the probability that no sample in the pool has a value higher than X^* :

$$P(\{X_i\} \leq X^*) = \prod_{i=1}^N \int_0^{X^*} dX = (X^*)^N. \quad (4.28)$$

Therefore, the probability that the sample with the highest prior mass has $X_N = X^*$ is

$$P(X_N = X^*) = \frac{\partial P(\{X_i\} \leq X^*)}{\partial X^*} = N X^{*N-1}. \quad (4.29)$$

A priori, X^* is an unknown bound, and we sample the prior masses from $0 < X < \tilde{X}$. For each X , one can define $s = X/\tilde{X}$, the *shrinkage factor*, which follows the same probability distribution as the highest prior mass

$$p(s) = N s^{(N-1)}. \quad (4.30)$$

So, once one has the highest prior mass \tilde{X} , it can be stored as a dead point, and the next point to add in the pool is $X = s\tilde{X}$, where s is directly sampled from Eq. (4.30). This leads to a series of prior masses

$$1, s_1, s_1 s_2, s_1 s_2 s_3, \dots \quad (4.31)$$

Since we have ensured that the next prior mass is smaller, the next point will also have a higher likelihood. We still explore the parameter space by increasing likelihood values. However, now, we do not need to re-order the prior mass for each sample anymore. One can then directly obtain the evidence by integration over the likelihood and prior masses of the retained points. For N dead points, we have

$$\mathcal{Z} \approx \sum_{i=1}^N w_i L_i, \quad (4.32)$$

where L_i is the likelihood associated with each point, and w_i is a measure of the distance between the consecutive prior masses. For example, one can simply take $w_i = X_i - X_{i+1}$ with $X_{N+1} = 0$ or use the trapezoid integration: $w_i = (X_{i-1} - X_{i+1})/2$ with the boundary conditions $X_0 = 2 - X_1$ and $X_{N+1} = -X_N$. Finally, regardless of w_i , Eq. (4.32) is a relatively simple expression and can be updated easily for each new dead sample added

to the queue.

In principle, one can continue the sampling process as long as one wants. There is no generic way to decide when the sampling is finished, and one needs to choose a *termination condition*. The latter should be a balance between computation time and precision on the final results. For example, one can decide to finish after a fixed number of iterations. However, the precision obtained after a given number of steps depends on the randomly taken shrinking factors. Therefore, this is not the best approach. Another approach is to decide to stop the sampling procedure when a new step does not bring a significant fractional change in the evidence anymore.

In nested sampling, in addition to the evidence, the posteriors are obtained “for free”. The output of the procedure outlined above is a set of samples $\{\boldsymbol{\theta}\}$, their corresponding likelihoods $\{\mathbf{L}\}$, and prior masses $\{\mathbf{X}\}$. From this, one can compute

$$W_i = \frac{L(\boldsymbol{\theta}_i)w_i}{\mathcal{Z}}, \quad (4.33)$$

the weight associated with each sample in the final posterior.

Since the nested sampling process has some probabilistic components, one can wonder about the method’s precision. Based on Eq. (4.30), it is possible to compute the mean and variance of the shrinking factors [147]:

$$\langle s \rangle = \frac{N}{N+1}, \quad (4.34)$$

$$\sigma_s^2 = \frac{N}{(N+1)^2(N+2)}. \quad (4.35)$$

Since $\sigma_s^2 \sim 1/N^2$, the variability of the factor is decreased by having more final points, hence sampling more before terminating. This reduces the variability and error in the evidence. Its statistical behavior is further characterized by its moments

$$\mathbb{E}[\mathcal{Z}^j]_{\mathbf{s}} = \int d\mathbf{s} (\mathcal{Z}(\mathbf{s}))^j P(\mathbf{s}), \quad (4.36)$$

where \mathbf{s} represents a series of shrinkage factors. If there is less variability in the shrinkage factor, then we should have more stable values in the integral, and therefore the first and second moments of the evidence should be more stable. These quantities are easy to compute as one can simply generate the series of shrinkage factors multiple times with a random number generator starting in a different initial state.

Variability in the process can also be induced by the starting step where we sample the prior. For a high-dimensionality problem, it may happen that the samples do not properly cover the initial space. Therefore, if one initializes the nested sampling algorithm multiple times, different outcomes can be obtained. One way to circumvent this effect is to run the algorithm several times in parallel and combine the resulting evidence and samples.

Finally, the exact number of initial points, termination condition, and other free parameters in nested sampling have to be chosen carefully depending on the problem considered. They should account for the dimensionality and complexity of the problem.

4.3.2 Nested Sampling for Gravitational Waves

Typically, for GW data analysis, one tries to compare the evidence under \mathcal{H}_{GW} (there is a GW in the data)

$$P(d|\mathcal{H}_{\text{GW}}) \propto \int d\boldsymbol{\theta} \mathcal{L}(\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{H}_{\text{GW}}), \quad (4.37)$$

where $\mathcal{L}(\boldsymbol{\theta})$ is the likelihood defined in Eq. (4.18), and under \mathcal{H}_{N} (there is only noise in the data)

$$P(d|\mathcal{H}_{\text{N}}) \propto \exp\left(-\frac{1}{2}\langle d|d \rangle\right), \quad (4.38)$$

where $\langle .|.\rangle$ is the noise-weighted inner product defined in Eq. (4.7). The ratio of the two then gives the Bayes factor expression of how likely it is to have a GW in the data³.

Publicly available tools have been developed to perform Bayesian analysis for GWs. The main frameworks used in the LVK collaboration are the C-based LALINFERENC package [148] and the python-based BILBY framework [149, 150]. In essence, the two encode the GW (4.18) and noise (4.38) likelihoods and are wrapped around a sampler which fixes most of the specifications required for the nested sampling process⁴. For example, possible samplers are DYNESTY [151], a *dynamic* nested sampling algorithm, where the number of live points is allowed to vary over time enabling to sample the posterior instead of having to make a prior transformation [152]. This sampler is the one used jointly with BILBY in the LVK parameter estimation runs [40]. As another example, we can mention the PYMULTINEST sampler [153], based on MULTINEST [154, 155]. The latter is a variant of nested sampling where the iso-likelihood contours are approximated by ellipsoids. More precisely, to also deal with multi-modalities, the samples are clustered and each cluster is represented by an ellipsoid.

4.4 Markov Chain Monte Carlo

Even if nested sampling is widely used in GW data analysis, other methods exist. One such method is the Markov chain Monte Carlo (MCMC) approach, where one evolves a set of samples taken from the prior through a series of transition functions – determining

³We assume the prior odds is unity here.

⁴Generally, they fix the initialization, termination condition, and exact mapping between prior masses and samples. Other parameters, such as the live points, are specified by the user.

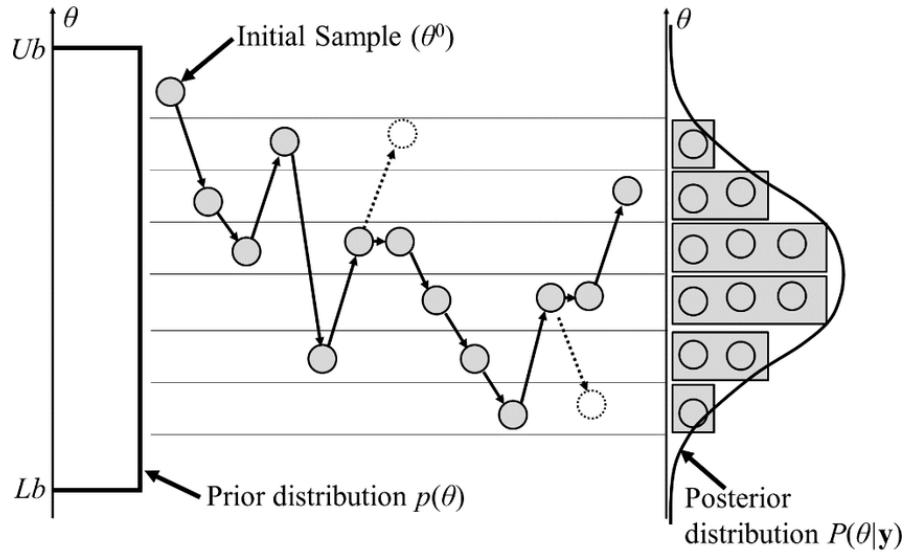


Figure 4.5: Illustration of the Metropolis-Hastings algorithm for a unidimensional problem. We start from a uniform distribution and take an initial sample. Then, one explores the space by applying the update process incorporating the proposal distribution depending on the data \mathbf{y} and a random number. The full circles are the points making it to the final chain, while the empty ones are discarded when computing their acceptance. We see that such points usually correspond to lower-density regions than the sample under consideration. In the end, the selected samples in the chain represent well the posterior distribution. Illustration adapted from [158].

what the next sample is – to get to the final posterior. At each step, one gets a sample θ which is added to the list of final samples, making a *chain* $\{\theta^0, \theta^1, \theta^2, \dots\}$ representing the posterior.

In particular, one often uses the *Metropolis-Hastings* algorithm [156, 157] to perform MCMC. Its principle is illustrated in Fig. 4.5. For the algorithm to work, one needs to choose a transition function determining the probability to go from a given sample to the next. Then, one builds the *acceptance ratio* A_r , which is the ratio of prior probabilities for the previous and new sample multiplied by the ratio of values for the proposal distribution when going from the previous to the new sample and vice-versa. If $A_r > 1$, one adds the new sample to the chain. When $A_r \leq 1$, one draws a random number R between 0 and 1. If $A_r > R$, the new sample is added to the chain; else, we add the old sample a second time in the chain. The process is repeated until we meet some convergence criterion.

Guided this way, the chain explores the parameter space and, provided we sampled long enough, provides samples corresponding to those one would have gotten by directly sampling the posterior distribution. While simple, this method has some drawbacks, and additional tricks need to be implemented. Some of the important techniques used are

- *Thinning*, where one removes some samples in the chain to remove the correlation

between successive samples [148] ;

- *Burn-in*, where one discards the samples obtained in the early stages. Indeed, since one starts with samples randomly taken from the prior, they do not necessarily represent the posterior and it takes several steps before the samples start being representative of the final distribution [148] ;
- *Parallel tempering*, where one introduces parallel chains with different likelihood temperatures, modifying $p(d|\boldsymbol{\theta}) \rightarrow p(d|\boldsymbol{\theta})^{1/T_i}$ for the i^{th} chain. This is a way to better explore the parameter space – avoiding the algorithm to focus only on one high-probability region when multiple exist – since the high-temperature chains have a smoothed likelihood, making it more homogeneous on the parameter space. Then, one swaps samples between the chains at a chosen rate. Eventually, important samples move to the lower temperature chains. In the end, the complex posteriors are well modeled by the chain with the lowest temperature [159] ;
- *Thermodynamic integration*, which enables one to compute the evidence when doing MCMC. Indeed, all the above has focused on getting accurate posteriors. However, to make a decision, one is also interested in the evidence. It is possible to show that the evidence is a function of the inverse chain temperature introduced in parallel tempering [160]. In this case, one can obtain the evidence by computing the expected log-likelihood for all the chains and integrating over the inverse temperature. In this case, the number of parallel chains influences the accuracy of the evidence, meaning one needs to choose their number carefully [148].

As for nested sampling, there exist routines one can use to perform MCMC analyses in the context of GW data analysis. LALINFERENCE [148] and BILBY (called BILBY-MCMC) [161] have MCMC capacities. The two code bases follow roughly the same approach, implementing the concepts introduced above. In particular, they rely on GW-specific proposals taking advantage of symmetries in the GW parameter estimation problem (for details, see [148, 161]), making the proposals more efficient, leading to faster convergence and a better parameter space coverage.

5

DEFLECTION OF WAVES BY MASSIVE OBJECTS

When a massive object is present on the travel path of a wave, it can deflect it, leading to *gravitational lensing*¹. This phenomenon is well-known for EM waves but can also happen for GWs. Einstein predicted it as a consequence of GR [162] but deemed it not observable because he considered lensing by stars. Later, this theory was put to the test for the first time by Eddington and Dyson, using a solar eclipse to show that the Sun's gravitational field leads to a deflection of background stars [4]. In 1937, Zwicky showed that one could have galaxies as lenses, leading to more common observations of the phenomenon [163]. However, before observing lensing, technological advancements were needed, and the observation of EM lensing became a reality in the 1970s [164].

Because its characteristics are useful to study different things, EM lensing has become a standard tool in astronomy. For example, it is used to discover exoplanets [5, 6]; when the planet passes in front of its host star, it will impact the observed light. Therefore, one can find the planet's mass and its distance to the host based on the observed (de)magnification of the host star. Another application is searching for dark matter [7–9]. Because the lensing effects depend on the total mass of the lens, measuring the magnitude of the time delay and magnification indicates the object's total mass. Comparing the inferred and visible masses, one can study the dark matter mass distribution. In addition, if one thinks about MACHO dark matter types, these objects should also lead to lensing effects. The frequency of lensing detections helps constrain the proportion of dark matter present in the Universe. Also, it can be used to probe the cosmological parameters [10–13]. In essence, by comparing the mass of the lens measured through the lens' geometry and the lens' dynamics, one has a relation linking two measurable values depending on the Hubble constant. So, putting the two equal enables one to compute its value. Finally, another interesting example is the study of distant objects [14, 15]. Thanks to its magnifying power, lensing allows observing objects that would otherwise be out of reach for telescopes, leading to more detailed studies of more distant objects.

The applications mentioned above focus on EM lensing. However, in the last years, GW lensing has been a growing research field, and many efforts are made to prepare for the first detection. For GWs, the principle is the same as for light: the wave gets deflected by a massive object present on its travel path [165–170]. The exact effect on the GW depends on the lens' characteristics. For lenses with a typical size smaller

¹In this work, we will often simply call it lensing.

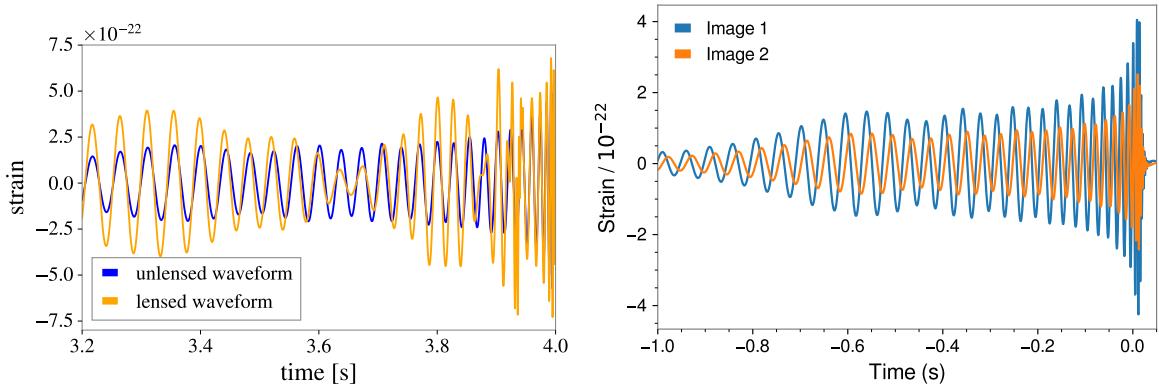


Figure 5.1: *Left:* Representation of the effect of microlensing on a GW waveform. It leads to frequency-dependent modulations. Figure taken from [190]. *Right:* Representation of the strong lensing effect on GWs. It leads to multiple images. These are magnified, shifted in time, and can undergo an overall phase shift. Here, the time delay is put to zero to see the phase difference. Figure taken from [191].

than the GW wavelength, one is in the *wave-optic limit*, and the waveform is distorted, leading to frequency-dependent beating patterns [168, 171–173]. A representation of this effect is given in the left panel of Fig. 5.1. These patterns depend directly on the lens' characteristics [174] and can therefore be used to study the objects present in the Universe, such as IMBHs [171, 175–177] or dark matter [178]. On the other hand, for larger lenses, we are in the *geometric optic limit* and one gets several images with the same frequency evolution [167, 168]. These images can be magnified, delayed in time, and undergo an overall phase shift [169, 170]. The magnification and time delay will mostly depend on the relative position of the source and the lens, as well as the lens' mass. For a given source-lens alignment, lower masses lead to superposed images and one faces *millilensing* [179]. On the other hand, for more massive lenses, the images are distinguishable. This is called *strong lensing* and is represented in the right panel of Fig. 5.1. Typically, strong lensing can occur due to a galaxy or a galaxy cluster lens. In the first case, the images can be separated from minutes to months [180, 181], while in the second case, they can be separated up to years [182–186]. In this thesis, we focus on strong lensing effects. The latter is predicted with an interesting rate in the coming years, with an increasing chance of observation as the detectors get upgraded [180, 187]. Therefore, searches have been ongoing since 2019 [172, 188], giving no confident detection but interesting candidates [189] requiring extensive analyses to determine their nature.

Since its observation is expected soon, it is of interest to know what benefits one could have from observing strong lensing. Therefore, it is of interest to know what benefits one could have from its observation. Many scientific applications rely on the virtually-extended detector network offered by the observation of multiple images. From the observation of one image to the other, the Earth undergoes a certain rotation and each image is seen from a different point of view. Combining the information obtained

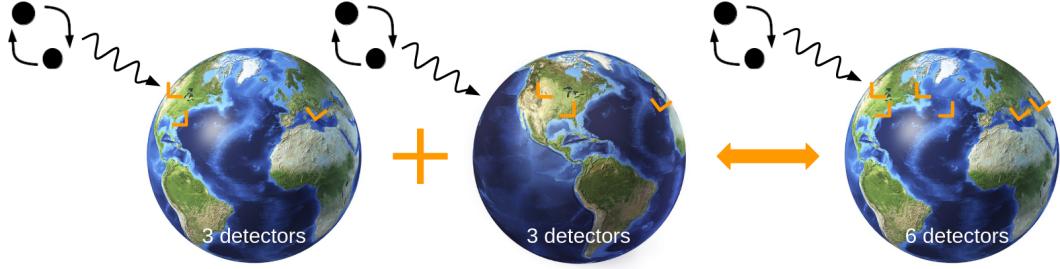


Figure 5.2: Illustration of the virtual extension of the GW detector network due to the detection of lensed images from the same event. In practice, the number of detectors for each image can vary due to downtime, changing the total number of detectors.

for each image, it is as if the event had been observed with a detector network made of the sum of detectors having observed each image (see Fig. 5.2 for an illustration when two images are detected). This leads to a more accurate sky location of the event [191–193]. Additionally, assuming the lensed binary is situated in a galaxy, the light emitted by the latter should also be lensed, and one can cross-correlate the EM and GW data [192, 193]. Once we have an observation in the EM and GW channels, one can perform precision cosmography [194–197], and test the speed of gravity [198, 199]. Even without an EM counterpart identification, strong lensing offers useful insights, such as the possibility to probe better the GW polarization content [200] and enhanced tests of GR in presence of HOMs. [117].

Currently, one of the main issues is the confident detection of strongly-lensed GW images. In principle, for a given set of detected BBHs, one analyzes all the possible pairs. It rapidly grows with the number of detection, leading to about 10^5 pairs to analyze when the LIGO and Virgo detectors reach their design sensitivity [180]. Therefore, the analyses need to be rapid. In essence, one looks for events with compatible characteristics. The probability of getting matches due to unlensed pairs increases as more BBHs are present in the pool of events [187, 201, 202]. Therefore, the analysis methods need also to be precise. Later in this work, GOLUM will be presented. This framework is a strong-lensing analysis pipeline solving some of the aforementioned issues. An extra way to decrease the risk of finding matches by chance is to constrain the lensing parameters according to some lens model [181, 187, 202, 203]. This can be accounted for in the search and data analysis processes, decreasing the FAR when the correct model is used [202].

In this chapter, we show how one can derive GW lensing. We then show the different lensing regimes and examples of lens models. Finally, we explain how one can search for lensing in GW data.

5.1 Gravitational-Wave Lensing

As presented in Sec. 2.2.1, one can see the GW as a perturbation on a background metric $\tilde{\eta}_{\alpha\beta}$. Since we are considering lensing, $\tilde{\eta}_{\alpha\beta}$ cannot be the usual Minkowski metric. So, the metric is the one corresponding to space-time with a gravitational potential $U(r) \ll 1$ due to the lens:

$$\tilde{\eta}_{\alpha\beta} dx^\alpha dx^\beta = -(1 + 2U)dt^2 + (1 - 2U)dr^2, \quad (5.1)$$

where the potential is assumed small because space-time is rigid, and one needs a very massive object before getting any distortion.

The GW leads to a perturbation $h_{\alpha\beta}$ of the background metric, leading to the total metric

$$g_{\alpha\beta} = \tilde{\eta}_{\alpha\beta} + h_{\alpha\beta}. \quad (5.2)$$

Similarly to Sec. 2.2.1, we place ourselves in the TT Lorentz Gauge. In addition, since we are only feeling the potential of the object and there is no matter, the EFEs (2.8) are [168]

$$\square h_{\alpha\beta} + 2\tilde{R}_{\alpha\beta\delta\gamma}h^{\alpha\delta} = 0, \quad (5.3)$$

where \tilde{R} is the Riemann tensor computed for the background metric (5.1).

Usually, the GW wavelength is much smaller than the radius of curvature. Therefore, the second term in Eq. (5.3) can be neglected, and we focus on the equation

$$\square h_{\alpha\beta} = 0. \quad (5.4)$$

The same type of equation was found when looking for plane wave solutions in Sec. 2.2.2. However, here, the d'Alembertian operator also contains information on the background metric (5.1), leading to a different solution.

It is convenient to decompose the GW into a scalar and a tensor component, called the *eikonal approximation* [204]

$$h_{\alpha\beta} = \chi t_{\alpha\beta}, \quad (5.5)$$

where $t_{\alpha\beta}$ is the tensor polarization and χ is a scalar. $t_{\alpha\beta}$ changes with an amplitude of the order of $U(r)$. Since the potential is assumed to be small, the change in tensor polarization is also small and can be neglected. Therefore, only the scalar component undergoes a variation, and we get the equation [168]

$$\partial_\alpha \left(\sqrt{-\tilde{\eta}} \tilde{\eta}^{\alpha\beta} \partial_\beta \chi \right) = 0. \quad (5.6)$$

Using Eq. (5.1) to express $\tilde{\eta}$ and taking the Fourier transform to get in the frequency domain, Eq. (5.6) becomes

$$(\nabla^2 + \omega^2) \tilde{\chi}(f) = 4\omega^2 U \tilde{\chi}(f), \quad (5.7)$$

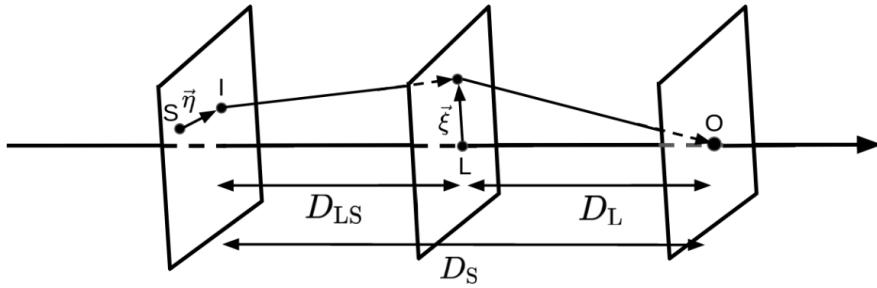


Figure 5.3: Representation of the setup to derive the amplification factor for GW lensing. Because of the lensing effect, the source appears displaced by a vector $\vec{\eta}$. We assume a thin lens with a straight propagation from the image to the lens plane. The GW crosses it at a position defined by $\vec{\xi}$ compared to the lens. It gets deflected in the lens plane and then continues in a straight line to the observer. The distance between the source and lens plane is D_{LS} , between the lens and the observer D_L , and between the observer and the source D_S . When accounting for cosmology, these distances are the angular diameter distances.

where the tilde expresses the transformation to the frequency space and $\omega = 2\pi f$. This equation can be solved using the Kirchhoff integral theorem [164].

To describe the modification due to lensing, it is easier to define the *amplification factor*

$$F(f) = \frac{\tilde{\chi}^L(f)}{\tilde{\chi}^U(f)}, \quad (5.8)$$

where $\tilde{\chi}^L$ and $\tilde{\chi}^U$ are the lensed and unlensed GW frequency-domain scalar amplitudes. $\tilde{\chi}^U$ is obtained simply by putting $U = 0$ in Eq. (5.7).

We now place ourselves in the thin-lens approximation. The image appears displaced relative to the source by a vector $\vec{\eta}$, and the GW crosses the lens plane at a position $\vec{\xi}$ compared to the lens. We assume the wave travels in a straight line between the planes and gets instantaneously deflected in the lens plane. The distance between the source and the observer, the source and the lens, and the lens and the observer are noted D_S , D_{LS} , and D_L , respectively. Fig. 5.3 shows the situation under consideration. In this case, one can show the amplification factor to be [164, 168]

$$F(f) = \frac{\xi_0^2 D_S}{D_L D_{LS}} \frac{f}{i} \int d^2 \vec{x} e^{2i\pi f t_d(\vec{x}, \vec{y})}, \quad (5.9)$$

where $\vec{x} = \vec{\xi}/\xi_0$, $\vec{y} = \vec{\eta} D_L / D_S \xi_0$ is the source position, and $t_d(\vec{x}, \vec{y})$ is the time delay to go from source to observer due to lensing. ξ_0 is a normalization constant. Finally, $F(f)$ is also normalized so that $|F| = 1$.

Because GWs can propagate through very long distances, cosmology needs to be included. Therefore, we need to adapt two quantities in Eq. (5.9). First, the distances need to be angular diameter distances. Second, we need to account for the frequency-redshifting provoked by the Universe's expansion. Therefore, the amplification factor (5.9)

becomes [204]

$$F(f) = \frac{\xi_0^2 D_S}{D_L D_{LS}} \frac{(1+z_L)f}{i} \int d^2 \vec{x} e^{2i\pi f t_d(\vec{x}, \vec{y})}, \quad (5.10)$$

where D_S , D_L , D_{LS} are now angular diameter distances, and z_L is the lens redshift. The arrival time t_d when going from $\vec{\eta}$ to the observer and passing through $\vec{\xi}$ is [164, 168]

$$t_d(\vec{x}, \vec{y}) = \frac{\xi_0^2 D_S}{D_L D_{LS}} (1+z_L) \left[\frac{1}{2} |\vec{x} - \vec{y}|^2 - \psi(\vec{x}) - \phi_m(\vec{y}) \right], \quad (5.11)$$

where the non-dimensional deflection potential $\psi(\vec{x})$ is found via

$$\nabla_x^2 \psi(\vec{x}) = \frac{\Sigma}{\Sigma_{cr}}, \quad (5.12)$$

where ∇_x^2 is the Laplacian in the x-direction, Σ is the surface density, and $\Sigma_{cr} = D_S/(4\pi D_{LS} D_L)$ is the critical mass density. $\phi_m(\vec{y})$ is the phase of the minimum time delay and can be chosen to set the minimum arrival time to zero. The exact expression for $\psi(\vec{x})$ and $\phi_m(\vec{y})$ depend on the lens. A few lens models will be presented in Sec. 5.2.

The amplification factor is constructed to link the lensed and unlensed waveforms. A lensed waveform $h_L(f)$ is related to the original unlensed one $h_U(f)$ in the frequency domain as

$$h_L(f) = F(f)h_U(f), \quad (5.13)$$

where $F(f)$ can only be described if we know the time delay, hence $\psi(\vec{x})$ and $\phi_m(\vec{y})$. The possible differences in scales between the lens and the GW lead to distinct regimes with different observable features, as we now discuss.

5.1.1 Geometric Optics Limit

When $f_{\text{GW}} \gg 1/t_d$, we are in the *geometrical optic limits*. In this case, Eq. (5.11) can be solved using the SPA described in Sec. 3.1. Only the stationary points of the time delay are contributing in Eq. (5.10). This can lead to several images, each with a position \vec{x}_j solution of [164, 168]

$$\frac{\partial t_d(\vec{x}, \vec{y})}{\partial \vec{x}} = 0. \quad (5.14)$$

Then, the amplification factor is

$$F(f) = \sum_j \sqrt{\left| \frac{1}{\det(\partial \vec{y} / \partial \vec{x}_j)} \right|} e^{(2i\pi f t_d^j - i\pi n_j)} = \sum_j \sqrt{|\mu_j|} e^{(2i\pi f t_d^j - i\pi n_j)}, \quad (5.15)$$

where μ_j is the magnification factor, translating the change in amplitude of the image compared to the unlensed waveform, t_d^j is the time delay for the j^{th} image, and $n_j = \{0, 0.5, 1\}$ if \vec{x}_j corresponds to a minimum, a saddle point or a maximum of the time delay, respectively. The magnification factor can be positive or negative and its sign determines the so-called *parity* of the image. Images with a negative parity – *i.e.* $\mu < 0$ – are mirror-symmetric images of the source [164].

So, for the geometrical optics limit, lensing can lead to several images, and their characteristics depend on the lens. If the time delay between the images is large enough, one has distinct images, and one talks about *strong lensing*. If the images overlap, one talks about *millilensing*.

In Eq. (5.15), one see a particular case arising when $\det(\partial\vec{y}/\partial\vec{x}_j) \rightarrow 0$, and $\mu_j \rightarrow \infty$. This is a non-physical result, meaning we get an infinite image magnification. The points for which this condition is met form the *critical* curves in the image plane. For these curves, one does not have an infinite magnification but a very large one. Therefore, the images are spread over a large area, meaning they are no longer a (stationary) point, and the approximation breaks down. By tracing the critical curves from the image to the source plane, one can find the *caustic* curves. The shape of these curves depends on the lens' geometry. The position of the source relative to the caustic has a major impact on the observed phenomenon. If the source is outside the curve, we generally see only one image. On the other hand, when it is inside, the image properties change, and we can see several of them with different properties [164], with, in general, a change of ± 2 images from one side of the caustic to the other. The critical curves also divide the images into different categories, with positive and negative parity. Depending on the position compared to the critical curves, the image magnification can be positive – the image has a positive parity – or negative – the image has a negative parity –.

5.1.2 Wave Optics

When the geometrical optics limit does not apply, the full integral in Eq. (5.10) needs to be solved. Typically, this leads to frequency-dependent beating patterns in the GW, hence we get a deformation of the wave with only one image formed. This often is referred to as *microlensing*. Another possibility is that we get wave optic effects in one of the images formed by strong lensing. This can happen when one of the images generated by a massive lens passes close to a smaller lens (*e.g.* a star in a galaxy) and gets an additional microlensing effect [178, 190, 205].

5.2 Lens Models

In the previous section, we showed the general effect of GW lensing on waveforms. However, the amplification factor (5.10) and the time delay (5.11) depend on the lens. In this section, we present several lens models. There exist many other lens models, and the list here is adapted to present models used in this thesis.

5.2.1 Point Mass Lens

First, we view the lens as a point mass. It is a simplified model, giving a first intuition on lensing. The lens' surface density is $\Sigma = M_L \delta^2(\vec{x}_i)$, with M_L the mass of the lens, and we take $\xi_0 = \sqrt{4M_L D_L D_{LS}/D_S}$ as normalization constant, and the non-dimensional deflection potential, found by solving Eq. (5.12), is $\psi(\vec{x}) = \ln(x)$. Under these conditions, Eq. (5.10) can be solved analytically, yielding [206]

$$F(f) = \exp \left[\frac{\pi w}{4} + i \frac{w}{2} \left(\ln(w/2) - 2\phi_m(y) \right) \right] \Gamma(1 - iw/2) {}_1F_1 \left(\frac{iw}{2}, 1; \frac{iwy^2}{2} \right), \quad (5.16)$$

where $w = 8\pi M_{Lz} f$, $\phi_m(y) = (x_m - y)^2/2 - \ln(x_m)$, $x_m = (y + \sqrt{y^2 + 4})/2$, $M_{Lz} = (1 + z)M_L$ is the redshifted lens mass, and ${}_1F_1$ is the confluent hypergeometric function. One already sees that the effect depends on two quantities: the redshifted mass of the lens and the position of the source compared to the observer-lens axis.

In the **geometric optics limit**, the amplification factor simplifies to [168]

$$F(f) = \sqrt{|\mu_+|} - i\sqrt{|\mu_-|} e^{2\pi i f \Delta t_d}, \quad (5.17)$$

with

$$\mu_{\pm} = \frac{1}{2} \pm \frac{y^2 + 2}{2y\sqrt{y^2 + 4}}. \quad (5.18)$$

The time delay between the two images is

$$\Delta t_d = 4M_{Lz} \left[\frac{4\sqrt{y^2 + 4}}{2} + \ln \left(\frac{y + \sqrt{y^2 + 4}}{\sqrt{y^2 + 4} - y} \right) \right]. \quad (5.19)$$

Here, $\Delta t_d \propto M_{Lz}$, and more massive lenses lead to longer time delays. In addition, $|\mu_{\pm}| \rightarrow \infty$ if $y \rightarrow 0$, meaning that if the source is closer to the observer-lens axis, one gets a larger magnification.

On the other hand, for the **wave optics limit**, one has to evaluate Eq. (5.16) numerically for each frequency.

5.2.2 Singular Isothermal Sphere

While the point mass lens is informative, it is definitively not so realistic. A more realistic model is that of a *singular isothermal sphere* (SIS) [168, 207], where the lens is a sphere with a characteristic constant velocity dispersion v . So, the surface density is $\Sigma(\vec{\xi}) = v^2/(2\xi)$. The normalization constant is taken as $\xi_0 = 4\pi v^2 D_L D_{LS}/D_{LS}$, and from Eq. (5.12) it follows that the non-dimensional potential is $\psi(\vec{x}) = x$. Under these conditions, the amplification factor takes the form [168]

$$F(f) = -iwe^{\frac{iwy^2}{2}} \int_0^\infty dx x J_0(wxy) e^{iw\left(\frac{1}{2}x^2 - x + \phi_m(y)\right)}, \quad (5.20)$$

where J_0 is the zeroth-order Bessel function, $\phi_m(y) = y + 1/2$ and $w = 8\pi M_{Lz}f$. Here, $M_{Lz} = 4\pi^2 v^4 (1 + z_L) D_L D_{LS}/D_S$ is the mass inside what is called the Einstein Radius².

For the **wave optics limit**, Eq. (5.20) has to be solved numerically. On the other hand, for the **geometric optics limit**, the amplification factor is given by [168]

$$F(f) = \begin{cases} \sqrt{|\mu_+|} - i\sqrt{|\mu_-|} e^{2i\pi d\Delta t_d} & \text{if } y \leq 1 \\ \sqrt{|\mu_+|} & \text{if } y \geq 1, \end{cases} \quad (5.21)$$

where $\mu_\pm = \pm 1 - (1/y)$ and $\Delta t_d = 8M_{Lz}y$. So, we get multiple images only if $y \leq 1$. In this case, the time delay between the images is still proportional to the mass of the lens as was the case for the point-mass lens model. If $y \geq 1$, only one magnified image appears.

5.2.3 Singular Isothermal Ellipsoid

Instead of describing a lens as a sphere, one can also use an ellipsoid with cylindrical symmetry. It is a more realistic description when modeling a galaxy since its rotation flattens the sphere one would see at rest. In this case, one needs to add a parameter to the SIS model: the axis ratio q , leading to a singular isothermal ellipsoid (SIE) [168, 208, 209]. The surface density for this lens model is $\Sigma(\vec{\xi}) = \sqrt{qv^2}/2\vec{\xi}$ [208]. This brings a second axis of symmetry to the problem, and we can have up to four images.

Unfortunately, this model is unsolvable analytically (both for the geometrical and wave optics limit), and one has to deal with the amplification factor and the lensing equation numerically. However, we can give a qualitative behavior depending on the

²The Einstein radius is a characteristic scale in lensing, corresponding to the radius of the ring appearing when there is a perfect alignment between the source, lens, and observer [164].

source alignment with respect to the observer-lens axis. When the source is on the outer limit, only one image is formed and has a relatively small magnification. It also has a positive parity. When the source gets closer to the caustic, a second image forms. This image has a negative parity and is usually demagnified. Finally, once the image is inside the caustic, two more images appear. One has a positive parity, and the other has a negative one. The magnification is very high at first when we are close to the caustic but decreases rapidly as the source gets closer to the central axis.

Finally, a still more complicated model, further breaking the symmetry of the problem, can be obtained by adding shear to the model. In this case, the axisymmetry is invalidated, and we typically get an odd pair of images. Such lens models require a numerical lens-equation solver, such as LENSTRONOMY [210, 211], for example.

5.3 Searching for Strongly-Lensed Events in the Data

Once we know what lensing does to the GW, one can look for it in the data. Such searches have been going on for some years in the LVK [172, 188], with no confident detection. In essence, it is a Bayesian inference problem, where one wants to determine whether we are in the lensed (\mathcal{H}_L , the GW data is lensed) or in the unlensed (\mathcal{H}_U , the GW data is unlensed) hypothesis, which can be assessed using the odds ratio (see Eq. (4.23)). The likelihood depends on the lens scenario since the lensing effect is different for different regimes.

Strong-lensing searches correspond to hypothesis testing (see Sec. 4.2.3 for an introduction), and we can distinguish two cases: we detect one or multiple images. From Eq. (5.15), one sees that the lensed and unlensed images are related via

$$h_L(f; \boldsymbol{\theta}, \boldsymbol{\phi}) = \sqrt{|\mu|} h_U(f; \boldsymbol{\theta}) e^{2\pi i f t_d - i\pi n}, \quad (5.22)$$

where $\boldsymbol{\theta}$ represents the usual BBH parameters and $\boldsymbol{\phi}$ the lensed parameters, hence the magnification, time delay, and Morse factor. However, generally, one has $h_U(f; \boldsymbol{\theta}) \propto 1/D_L$, where D_L is the luminosity distance. Therefore, there is a degeneracy between the magnification and the luminosity distance. So, an image is observed with an *apparent* luminosity distance

$$D_L^{\text{app}} = \frac{D_L}{\sqrt{|\mu|}}. \quad (5.23)$$

Moreover, the lensing time delay term is indistinguishable from the unlensed coalescence time t_c . Therefore, we have an apparent time of coalescence

$$t_c^{\text{app}} = t_c + t_d. \quad (5.24)$$

Finally, while it can look like the additional phase due to the Morse factor n_j can be

absorbed in the coalescence phase, it is not the case in all generality. Looking at a waveform with HOMs (3.27), the phase of coalescence appears with different exponents in different modes, avoiding the degeneracy with the coalescence phase. Therefore, in principle, one could detect this effect due to lensing when observing a **single image**. However, this effect is present only when one observes a so-called type II image ($n_j = 0.5$). For other values, the effect corresponds to no shift (for $n_j = 0$, called type I), or a mirror shift ($n_j = 1$, called type III), and no distortion can be seen [170]. Fig. 5.4 shows a comparison between a type I and a type II image. Shifting the unlensed waveform is not helping to recover the correct signal, and additional, potentially detectable, deformations are present. Mathematically, it can be understood as follows. First, one can re-express the spherical harmonic decomposition (3.27) in the time domain as [212]:

$$h(t) = \sum_{l \geq 2} \sum_{m \geq 0} A_{lm}(t) \cos(\phi_{lm}(t) + m\phi_c - \chi_{lm}), \quad (5.25)$$

where A_{lm} is an amplitude term depending on the binary's intrinsic parameters, its sky location (α, δ) , the inclination ι and the polarization ψ , ϕ_c is the phase of coalescence, and

$$\chi_{lm} = \tan^{-1} \left(\frac{F_{\times}(\alpha, \delta, \psi) f_{\times}(\iota)}{F_{+}(\alpha, \delta, \psi) f_{+}(\iota)} \right), \quad (5.26)$$

with $f_{\times,+}(\iota)$ a function representing the dependence on the inclination for each polarization, and $F_{\times,+}$ are the detector's antenna response (see Eq. (2.26)).

Combining the strong lensing effect (5.22) with Eq. (5.25), one sees that the waveform for the different image types are³:

- $h_{\text{I}} = |\mu_{\text{I}}| \sum_{l \geq 2} \sum_{m \geq 0} A_{lm}(t) \cos(\phi_{lm}(t) + m\phi_c - \chi_{lm}).$

So, it corresponds to a GR waveform with an overall magnification.

- $h_{\text{II}} = |\mu_{\text{II}}| \sum_{l \geq 2} \sum_{m \geq 0} A_{lm}(t) \cos(\phi_{lm}(t) + m\phi_c - \chi_{lm} + \pi/2).$

Since the phase of coalescence appears with a different pre-factor for each mode, it is not degenerate with the Morse factor. So, the latter leads to deformations in the waveform and can be detected – see Fig. 5.4 for a representation.

- $h_{\text{III}} = |\mu_{\text{III}}| \sum_{l \geq 2} \sum_{m \geq 0} A_{lm}(t) \cos(\phi_{lm}(t) + m\phi_c - \chi_{lm} + \pi).$

The extra π shift here corresponds to a global sign flip of the wave. It can be shown [213] that the complex waveform can be written $h(f) \sim A'(f) e^{2i\psi}$. Therefore, the global sign change can be mimicked by changing the polarization angle as $\psi \rightarrow \psi + \pi/2$, and single type III images cannot be identified.

³While not written explicitly here, the time in waveforms correspond to the lensed time, hence $t = t_{\text{U}} + \delta t_{\text{L}}$.

Therefore, one can assess how likely it is for a given GW event to be a lensed type II image as follows. Writing the strain

$$d(f) = n(f) + h_{\text{H}}(f), \quad (5.27)$$

where n is the noise component and h_{H} is a GW which can be unlensed ($h_{\text{H}} = h_{\text{U}}$) or lensed ($h_{\text{H}} = h_{\text{SL}}$). Assuming the prior odds does not play an important role, we can evaluate which hypothesis is the most likely through the Bayes factor

$$\mathcal{B}_{\text{U}}^{\text{SL}} = \frac{p(d|\mathcal{H}_{\text{SL}})}{p(d|\mathcal{H}_{\text{U}})} = \frac{\mathcal{Z}_{\text{SL}}}{\mathcal{Z}_{\text{U}}}, \quad (5.28)$$

where \mathcal{Z}_{SL} and \mathcal{Z}_{U} are the evidence under the lensed and unlensed hypotheses.

Under the unlensed hypothesis, the GW is described by its usual parameters $\boldsymbol{\theta}$ ($h(f) = h_{\text{U}}(f; \boldsymbol{\theta})$) and the evidence is

$$\mathcal{Z}_{\text{U}} = \int d\boldsymbol{\theta} p(d|\boldsymbol{\theta}, \mathcal{H}_{\text{U}}) p(\boldsymbol{\theta}|\mathcal{H}_{\text{U}}), \quad (5.29)$$

where we use the likelihood and prior as explained in Sec 4.2.1.

Under the strongly-lensed hypothesis, the waveform is described by $h_{\text{SL}}(f) = h(f; \boldsymbol{\vartheta}) e^{-i\pi n \text{sign}(f)}$, where $\boldsymbol{\vartheta}$ are the parameters describing the unlensed waveform with the apparent luminosity distance and coalescence time, and n is the Morse factor. Under these conditions, the evidence is

$$\mathcal{Z}_{\text{SL}} = \int d\boldsymbol{\vartheta} dn p(d|\boldsymbol{\vartheta}, n, \mathcal{H}_{\text{SL}}) p(\boldsymbol{\vartheta}, n, |\mathcal{H}_{\text{SL}}). \quad (5.30)$$

The Bayes factor (5.28) can then directly be computed by taking the ratio of Eqs. (5.29) and (5.30). If $\ln(\mathcal{B}_{\text{U}}^{\text{SL}})$ is significantly larger than zero, then the event is a type II strongly-lensed image, leading to smoking-gun evidence for strong lensing. The main drawback of this approach is that one needs very large HOMs to detect the effect, requiring particular BBH to be lensed or very large SNRs [115, 116].

When **multiple images** are present, the hypothesis framework is slightly different since one has to accommodate the different images. Here we outline how it works when comparing two data streams, hence assuming we observe two potentially lensed images. This can easily be adapted to more images. In this context, one observes two data streams d_1 and d_2 , each made of a GW and a noise component. So,

$$d_i = n_i + h_i(\boldsymbol{\Xi}_i, \mathcal{H}), \quad (5.31)$$

where $i = \{1, 2\}$, and $(\boldsymbol{\Xi}_i, \mathcal{H}) = (\boldsymbol{\theta}_i, \mathcal{H}_{\text{U}})$ if we are in the unlensed hypothesis, and $(\boldsymbol{\Xi}_i, \mathcal{H}) = (\boldsymbol{\theta}, \boldsymbol{\phi}_i, \mathcal{H}_{\text{L}})$ if we are in the lensed one. So, under the unlensed hypothesis, the two images are independent and have different parameters, while for the lensed hypothesis, we have the same BBH parameters but linked through the lensing parameters,

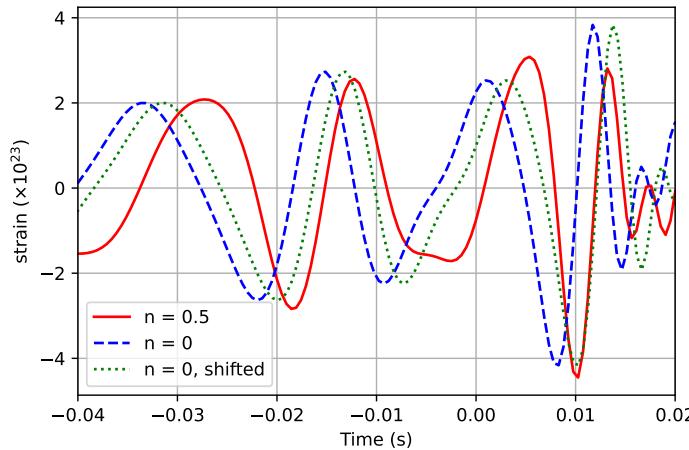


Figure 5.4: Representation of type I and type II images for an event with large HOM contributions. Shifting the type I image does not make for a waveform corresponding to a type II image, meaning we cannot simply reproduce the type II image by fine-tuning the usual CBC parameters. Therefore, it is a potentially observable effect.

different for each image. One is then interested in the ratio of evidence

$$\mathcal{C}_U^L = \frac{p(d_1, d_2 | \mathcal{H}_L)}{p(d_1, d_2 | \mathcal{H}_U)}, \quad (5.32)$$

usually called the *coherence ratio* in the lensing framework. One talks about a *Bayes factor* when selection effects are accounted for [214]. We will talk more about it in Part II of this thesis, where we show the development of strong-lensing searches.

First, for the unlensed hypothesis, the evidence

$$\begin{aligned} p(d_1, d_2 | \mathcal{H}_U) &= \int d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2 p(d_1, d_2 | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \mathcal{H}_U) p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 | \mathcal{H}_U) \\ &= \int d\boldsymbol{\theta}_1 p(d_1 | \boldsymbol{\theta}_1) p(\boldsymbol{\theta}_1 | \mathcal{H}_U) \int d\boldsymbol{\theta}_2 p(d_2 | \boldsymbol{\theta}_2) p(\boldsymbol{\theta}_2 | \mathcal{H}_U) \\ &= p(d_1 | \mathcal{H}_U) p(d_2 | \mathcal{H}_U), \end{aligned} \quad (5.33)$$

where we used the fact that unlensed events are independent. In the end, the unlensed evidence for two data streams d_1 and d_2 is the product of the individual evidence for each data stream. If one has more than two images, the conclusion remains valid, and the unlensed evidence is the product of the individual evidence under the unlensed hypothesis.

Under the lensed hypothesis, one needs to account for the dependency between the

images. Therefore, the evidence

$$\begin{aligned} p(d_1, d_2 | \mathcal{H}_L) &= \int d\boldsymbol{\theta} d\boldsymbol{\phi}_1 d\boldsymbol{\phi}_2 p(d_1, d_2 | \boldsymbol{\theta}, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \mathcal{H}_L) p(\boldsymbol{\theta}, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, | \mathcal{H}_L) \\ &= \int d\boldsymbol{\theta} d\boldsymbol{\phi}_1 d\boldsymbol{\phi}_2 p(d_1 | \boldsymbol{\theta}, \boldsymbol{\phi}_1, \mathcal{H}_L) p(d_2 | \boldsymbol{\theta}, \boldsymbol{\phi}_2, \mathcal{H}_L) p(\boldsymbol{\theta}, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, | \mathcal{H}_L), \end{aligned} \quad (5.34)$$

where $\boldsymbol{\theta}$ represents the binary's parameters – common for the two images – and $\boldsymbol{\phi}_i$ the lensing parameters for image i . Under this form, the analysis has to take in the two data streams jointly to have the joint evidence for the lensed hypothesis [214, 215]. If more than two images are formed, then one needs to add an integration over an extra set of lensing parameters⁴.

Once Eqs. (5.33) and (5.34) have been evaluated, one can get the coherence ratio and see whether the event is lensed or not. Evaluating the \mathcal{C}_U^L using the full joint expression is computationally heavy. In this thesis, we will show how the likelihood can be recast as a conditional likelihood to allow for both speed and precision [191]. Another approach is the *posterior overlap* method [181]. Combining Eqs. (5.34) and (5.33), the coherence ratio, considering only the parameters unaffected by lensing ($\boldsymbol{\vartheta}$), can be expressed as [181]

$$\mathcal{C}_U^L = \int d\boldsymbol{\vartheta} \frac{p(\boldsymbol{\vartheta} | d_1) p(\boldsymbol{\vartheta} | d_2)}{p(\boldsymbol{\vartheta})}, \quad (5.35)$$

which is an integral computing the agreement between the posteriors. Typically, posterior overlap uses a sub-set of parameters and a kernel density estimator (KDE) to reconstruct them before evaluating Eq. (5.35) [181]. This method is much faster than the joint parameter estimation ones [214, 215] but it is also less precise since it does not account for the full correlation of all the parameters. Moreover, it does not analyze the signals jointly, making it unable to extract information on the lensing parameters. Nevertheless, it can be used as a first filtering step when searching for strong lensing.

Including a lens model can help reduce the false-alarm risk related to strong lensing [181, 187, 202, 203]. Its inclusion in the coherence ratio is detailed later in this thesis. Often, one simply complements the coherence ratio with the ratio of probabilities for the lensing parameter under lensed and unlensed hypotheses

$$\mathcal{S}_U^L = \frac{p(\boldsymbol{\phi} | \mathcal{H}_L)}{p(\boldsymbol{\phi} | \mathcal{H}_U)}, \quad (5.36)$$

where \mathcal{S} changes depending on the lens model used, and $\boldsymbol{\phi}$ can be a subset of or all the lensing parameters. The statistic can be based on different lens models (see Refs. [181, 187, 203] for example statistics based on different lens models and parameters). Generally, to compute these probabilities, one makes an unlensed BBH population and

⁴In this case, the joint inference of the multiple images can become computationally expensive, and practically intractable.

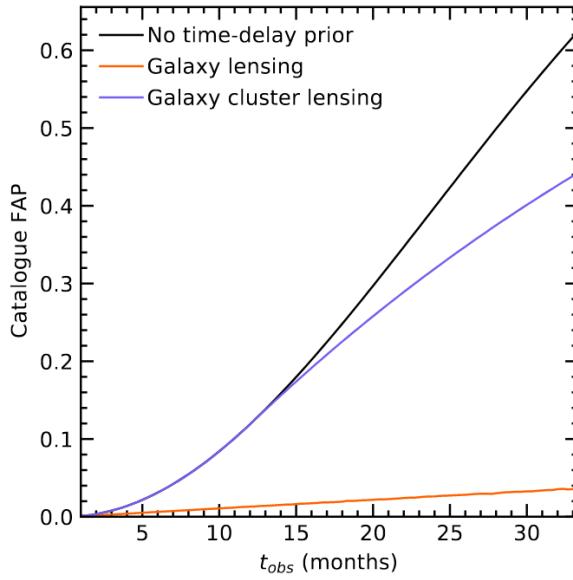


Figure 5.5: Representation of the FAP for lensed pairs as a function of the total observation time, with and without lens model. The inclusion of a model through a probability ratio decreases the FAP. The effect is more important for a galaxy-lens model than for a galaxy cluster lens because the overlap between the (apparent) lensing parameters for the lensed and unlensed population is lower for galaxy lenses. Figure adapted from [187].

computes their apparent lensing parameters⁵. For the lensed distributions, one makes a BBH and a lens population, assuming some merger rates and property distributions. One sees how many events are lensed and what are the lensing parameters. Once there are enough samples for the lensed and unlensed hypotheses, one can make a KDE reconstruction (for example) of the parameters and use this as a probability density function to evaluate Eq. (5.36). Accounting for this ratio in the detection statistic can help decrease the FAP [187, 202]. Fig. 5.5 represents this effect. One sees the inclusion of a model significantly reduces the FAP, especially for a galaxy lens model. The effect is weaker when considering a galaxy-cluster lens model because galaxy-cluster lenses lead to a larger diversity of time delays and magnifications. In particular, they often have longer time delays. Therefore, there is more overlap between lensed and unlensed probability distributions, and Eq. (5.36) is less effective [187].

⁵The observed relative magnification is the square root of the luminosity distance ratio, the time delay is the difference in coalescence time, and the apparent Morse factor difference can be computed through the difference in phase of coalescence.

6

MACHINE LEARNING CONCEPTS

Machine learning (ML), in particular, deep learning (DL), has seen major developments in the last few years. It is applied in a growing number of fields, thanks to its ability to deal with high dimensional data, its flexibility, and its speed after training. Some major successes for ML are its use in self-driving cars [216, 217], medical image analysis [218], image cleaning for astrophysical surveys [219–221], and many others (e.g. [222–225]). In particular, it is also more and more commonly used in GW data analysis; see Ref. [62] for an overview.

ML is a computational technique of learning through examples. In particular, DL is a sub-branch of ML where one uses multi-layered, non-linear structures – called *neural networks* (NNs) – to go from data to an output. They lead to more freedom in fitting the problem under consideration and better performances of the applied method. In this thesis, we focus on supervised learning methods, meaning that the algorithm is trained using labeled examples given by the user. Hence, there is a set of data with their true value given to the algorithm. The latter then adjusts its configuration through a learning process to improve its performance.

In this chapter, we describe the basics of supervised-learning ML algorithms as well as some specific techniques related to inference problems.

6.1 The Building Blocks

Originally, NNs were thought of as a sort of numerical representation of the human brain. Therefore, common terms are present between ML and neuroscience. The jargon is detailed in this section. In practice, the idea behind a NN is to have an algorithm capable of learning and adapting by seeing examples.

6.1.1 The fundamental Entity: the Neural Network

The main entity used in ML is a NN. It is a collection of *neurons*, making a multi-layered and non-linear structure. A neuron takes several inputs, combines them in a weighted

sum, and adds a bias. Mathematically,

$$\hat{n}(\vec{x}) = b + \sum_{j=1}^N w_j x_j, \quad (6.1)$$

where $\vec{x} = \{x_1, \dots, x_N\}$ is the input vector, $\{w_j\}_{j=1, \dots, N}$ is the set of weights, and b is the bias. The weights and bias form trainable parameters, adjusted through the training process.

Generally, one also applies an *activation function* f to the neuron's output

$$f(\vec{x}) = f(b + \sum_{j=1}^N w_j x_j). \quad (6.2)$$

The simple structure formed by a neuron and an activation function is called a *perceptron*, first described by Rosenblatt in 1958 [226]. It can be seen as a primitive form of a NN but is generally not complex enough to fulfill realistic tasks. NNs can also be seen as an ensemble of perceptrons. Two examples of typical activation functions are:

- the ReLU (Rectified Linear Unit):

$$\text{ReLU}(x) = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{else.} \end{cases} \quad (6.3)$$

It weights the values based on their sign, and generally leads to a rapid convergence for the network;

- the GELU (Gaussian Error Linear Unit):

$$\text{GELU}(x) = x / (1 - e^{-1.706x}). \quad (6.4)$$

It focuses on the input's value rather than only its sign. In addition, it does not suffer from “dying neurons” like the ReLU function, where once the derivative is set to zero, it remains zero for the rest of the process [227].

The activation functions are chosen by the user and should be adapted to the task at hand.

A NN is then a collection of neurons linked through activation functions and grouped as layers. For a given neuron in a given layer, its output passes through an activation function before being used as one of the inputs for the neurons in the next layer. This leads to more flexibility and higher representation capabilities when learning the problem. The number of neurons, hence the number of weights and biases, can change from one layer to the other. The collection of layers leads to a differentiable and non-linear structure, generally without an analytical solution. Its weights and biases are adjusted through a process called *training*.

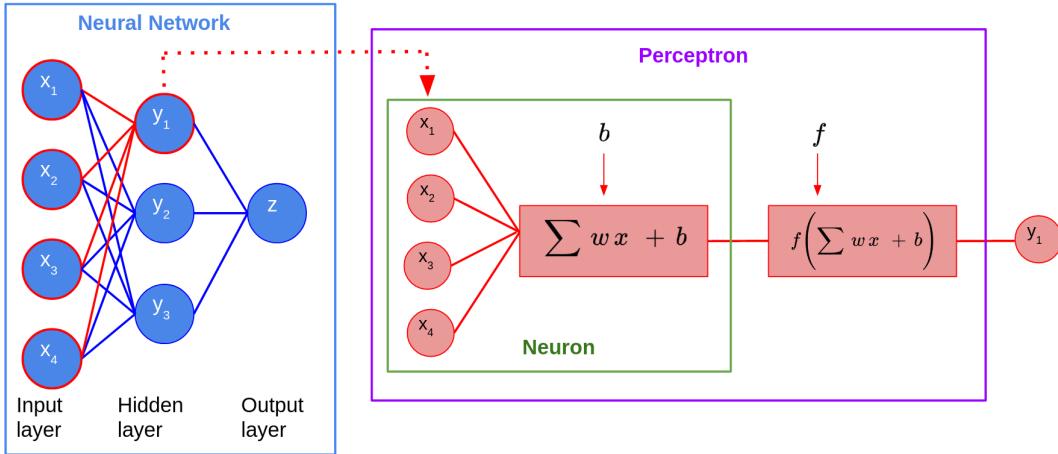


Figure 6.1: Illustration of a simple neural network, a perceptron, and a neuron with the link between them. In a neural network, the output of a neuron passes through an activation function before being used as input for the neurons of the next layer. A neuron followed by an activation function corresponds to a perceptron.

Typically, one distinguishes three layer types:

- the *input layer*: the set of neurons taking in the data. It can be raw data or a representation of it ;
- *hidden layers*: the next layers after the input one, except for the last layer. They represent different levels of abstraction, extracting different features of the data. Their number varies from one network to the other ;
- the *output layer*: the final layer, whose output is the sought information. For example, the class if we do classification, *i.e.* determining the class corresponding to the data, or a value if we do regression, *i.e.* determining the value of one or more parameters based on the data.

If the network has too few layers, it can be unable to model the situation, leading to poor performance. On the other hand, if it is too large, it can overfit, reducing flexibility. In this case, the NN learns the mapping from example to true value for the training data but is unable to generalize to data outside of the learning examples. A representation of a NN, a perceptron, and a neuron, as well as their interconnections, is given in Fig. 6.1.

6.1.2 Loss Functions, the Key to Training

To adjust the network's performance, one requires a metric to assess how close the NN's outputs are to the true values in the training data. The *loss function* is a way to measure this. It has to be adjusted to fit the task under consideration.

A simple example of a loss function is the mean-square error (MSE)

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (O_i - T_i)^2, \quad (6.5)$$

where O_i is the network's output and T_i is the true value for the example i . This can be used for simple regression problems but is often not suited for more realistic scenarios.

For classification tasks, one often uses the cross-entropy [228]. It is based on the notion of information in probabilities and measures the difference between two probability distributions for given sets of events. The cross-entropy between two probability distributions P and Q is

$$H(P, Q) = - \sum_{x \in X} P(x) \log(Q(x)), \quad (6.6)$$

where X is the set of samples for which the cross entropy is computed. When training a NN, $p(x)$ is the probability to be in a given class based on the true value, hence it is 1 for the correct class and zero for all the others. $Q(x)$ is then the probability to be in a given class according to the NN's output. Clearly, the loss is maximized when $Q(x) = 1$ for the correct class.

Eq. (6.6) resembles the Kullback-Leibler (KL) divergence, which describes how much two distributions differ from each other and can also be used as a loss function, even if it is not symmetric in the two distributions, making its use somewhat difficult:

$$KL(P, Q) = - \sum_{x \in X} P(x) \log \left(\frac{Q(x)}{P(x)} \right). \quad (6.7)$$

If the KL statistic is small, the two distributions are very similar, and the network outputs values close to the true ones.

As another example, let us also note that it is possible to use the negative Bayesian log-likelihood as a loss function if it is known for the situation. The loss function is minimized during the training (see Sec. 6.1.3 for more details). For the correct parameters, the log-likelihood is maximum. So, the negative log-likelihood is at its minimum. Therefore, we seek to find parameters that minimize the negative log-likelihood since they are a good representation of the data. In this case, one does not need a direct comparison with the true values making this approach more suited when the goal is to model probabilistic distributions.

6.1.3 Gradient Descent, the Key to a Working Network

Once a network is built, it has a certain number of weights and biases to adjust to minimize the loss function. The best performance in the training set is obtained when

we are at the minimum of the loss function. The latter can be seen as a function of the bias and weights, making a hyper-surface in their space. Ideally, we want to find the deepest valley on this surface, corresponding to the global minimum. However, we often have to be satisfied with a good approximation, corresponding to a local minimum. This is done through an iterative process called *gradient descent*.

Since the loss space is complicated, the trick is to consider only local information and simplify the function at that point of space. Therefore, we can do a Taylor expansion of the loss. Writing the network parameters ζ and the loss function L , its expansion around a given network configuration ζ_j is

$$L(\zeta_j + \epsilon) \simeq L(\zeta_j) + \epsilon \cdot \nabla L(\zeta_j) + \mathcal{O}(|\epsilon|^2), \quad (6.8)$$

where ϵ is a small step in the network parameter space, meaning its norm is small. The goal is to find a step leading to smaller values for the loss, hence an ϵ such that $L(\zeta_j + \epsilon) < L(\zeta_j)$, and which can be computed based on the loss. This is the case if we choose

$$\epsilon = -\gamma \nabla L(\zeta_j), \quad (6.9)$$

where γ is a (small) positive number that can be tuned during the training process. With this, the network parameters follow the update rule,

$$\zeta_{j+1} = \zeta_j - \gamma \nabla L(\zeta_j), \quad (6.10)$$

when going from a network configuration j to $j + 1$ when minimizing the loss function.

Using Eq. (6.9) as step, we have

$$L(\zeta_{j+1}) = L(\zeta_j + \epsilon) \simeq L(\zeta_j) - \gamma \nabla L(\zeta_j) \cdot \nabla L(\zeta_j) < L(\zeta_j), \quad (6.11)$$

since $\gamma \nabla L(\zeta_j) \cdot \nabla L(\zeta_j) > 0$. This shows that using Eq. (6.9) for the step, we move to lower loss function values at each iteration. So, when training the network and minimizing the loss function, one starts with some initialization of the weights. Then, one computes its derivative and uses it to move to a lower loss value. Once the step is done, the process is repeated until we reach a satisfactory minimum of the loss function. In this process, the free parameter γ determines the step size. It is called the *learning rate* and is chosen by the network's designer. It has to be chosen with care. If it is too large, we can over-shoot the minimum and miss it. If it is too small, the training takes a very long time, and there are risks to be stuck at a local minimum. Among other things, one can make this rate evolve, starting with a large value to quickly move to the interesting region of the parameter space before reducing it to find the correct overall minimum in that region.

6.1.4 Better Explore the Loss Function Space: Optimizers

Gradient descent, as described in the previous section, is the basic principle to obtain optimal network parameters. However, it is often inefficient, and the process is made more effective using an *optimizer*. A common optimizer family is the Adam family, based on the Adam optimizer [229], derived from the adaptive moment estimation.

The Adam algorithm uses moving averages to keep track of the mean and variance of the gradient for each network parameter. This information is used to grasp the loss's behavior better. In particular, it prevents missing a good minimum by keeping track of the dynamics. For example, if there is a small local minimum, then the momentum will push us out of it based on past dynamics. On the other hand, the loss will also not oscillate around a point, as the average over the oscillatory steps stabilizes the value of the loss to the actual minimum. Additionally, Adam is robust to noise in the gradient computation by averaging the values over several steps. This is a valuable feature since gradients are evaluated numerically, and one can have imprecise values due to the problem's high dimensionality and noise. In the end, these features combined with an efficient algorithm make Adam a popular optimizer with good performances for many ML tasks [229].

There are also variants of Adam designed to be more efficient or solve remaining issues. For example, there is the AdaMax, where instead of using the moving average of the mean and variance of the loss, we use the infinity norm of the gradient vector [229]. This modification improves the robustness against noise and sparse gradients where, in some cases, the gradient can take low values, making Adam inefficient. Therefore, taking the maximum value helps keep a more stable learning rate and solve the optimization problem more efficiently [229]. Another improvement is presented in the AdamW algorithm [230], where one also implements *weight decay regularization*. Here, one slightly decreases the value of the network parameters when updating them to new values. The simple update rule (6.10) is then modified as

$$\zeta_{j+1} = (1 - \lambda)\zeta_j - \gamma \nabla L(\zeta_j), \quad (6.12)$$

where λ is a small positive number. This additional decrease in parameter values can also be implemented in the moving averages used in the other Adam algorithms. λ can be seen as a penalty term for the network parameters. By decreasing their values, one avoids overfitting since weights that would become dominant are rendered less important. In parallel, unimportant weights are decreased more rapidly as the update rule increases their value less, and the penalty is not compensated. The value of λ can be kept fixed or adapted over time. Using this additional parameter in the update rule decreases the risks of overfitting, improving the network's generalization capacities [230].

6.1.5 Overview of a Basic Training Process

Now that we have outlined the different elements needed to build and train a NN, we can briefly sketch how supervised learning works. We start with a set of example data with matching true values. Typically, one splits this set in two: the *training* and *validation* sets. The training set is used to adapt the weights, while the validation set is used to gauge the evolution of the performance over time. It is also good practice to have a third set of labeled data, the *test* set, used only once to gauge the performance on unseen data after the training is stopped.

When training the network, one passes the training set through it, adjusting the weights as one goes. Then, one checks the results on the validation set. If the performances do not seem adequate yet, one continues. Else, one can stop the training and assess the performance on the test set. Often, one does not pass the whole training set at once, using batches instead. So, one adapts the weights based on the results of a batch of data. The complete passage of the batched training set is called a *training epoch*. The training termination condition can be set by specifying a given number of epochs. However, this is generally not optimal since the number of steps needed for convergence is unknown beforehand. Instead, one can decide to put a termination condition based on the evolution of the loss function or the performances on the validation set.

Stopping the training on time is also a key aspect. Indeed, the loss function can continue its decrease for the training set until it reaches very low values. However, at some point, there is no improvement in the validation set anymore. On the contrary, we get better performances on the training set, while the performances on the validation set degrade. This is called *over-fitting*. In this case, the weights and biases are not adjusted to features of the data anymore but rather to the training set itself. Therefore, the network loses its adaptation capabilities and cannot perform well on unseen data. One way to deal with this is to have much larger data sets. This is often difficult due to memory issues. Another way to circumvent this is to generate new data continuously while the training is ongoing. This is possible only with a fast data generation process.

Finally, let us outline an additional technique used for the training and helping the network to learn more complicated data: *curriculum learning*. Typically, there are easier and harder cases in the data. If we simply sample the data randomly, the network has a hard time interpreting the more difficult instances and performs poorly on them. To avoid this, one can gradually increase the difficulty of the data passed to the network when training. We start by showing simple examples. Once the network performs well on them, we move to the next more complicated data set. Usually, one keeps a fraction of the simpler examples in the next sets to avoid the network forgetting them. In principle, one expects the network to have satisfactory performance for all the sets in the data after the full training is done.

6.2 Finding Spatial and Time Correlation in Data

In the previous section, we introduced the main elements of a NN. We have only presented *dense* or *fully-connected* layers, connecting all the neurons of one layer to all the neurons of the next layer in a biased and weighted sum. Networks built only with such layers often have many free parameters and can be hard to train. Therefore, we introduce *convolutional neural networks* (CNNs), which alleviate some of these difficulties. Such networks are made to find correlations on different levels in the data by using *convolutions*. This type of NN is the subject of this section.

6.2.1 Convolutional Neural Networks

Instead of connecting one neuron to all the neurons of the next layer, the convolution connects a sub-group of neurons to the next neuron. The number of neurons considered is called the *kernel size*. The neurons are then summed and biased like before. The convolution with a given kernel side slides over the layer to collect information from the full layer. The step taken is called the *stride*. One can also add *padding* to compensate for a mismatch between the number of neurons required based on the stride and kernel size, and the number of them available in the layer. For example, one can add null neurons. The working of a convolutional layer is represented in Fig. 6.2. Because they look at local information in a given layer, convolutional layers contribute to translation invariance in the NN.

One can stack convolutional layers one after the other to build a CNN. Each layer probes a different correlation level in the data. In a CNN, one usually adds *pooling layers*, which are filters reducing the dimension of a layer by summarizing its content. If well applied, it can help increase the network's efficiency by helping it focus on the key feature. Besides, the combination of pooling and convolutional layers leads to translation invariance by bringing the key features into the NN's focus. However, they can also lead to a loss of information if applied too extremely. The number of neurons considered when doing an average operation is called the kernel size. We distinguish two typical pooling layers:

- *Max pooling*: one selects the maximum value present in the kernel under consideration and passes it to the next neuron ;
- *Average pooling*: one computes the average value present in the kernel and passes it to the next neuron.

A complete CNN is typically made by an input layer, taking in the data. Then, one stacks multiple convolution blocks, made by convolution layers followed by a pooling

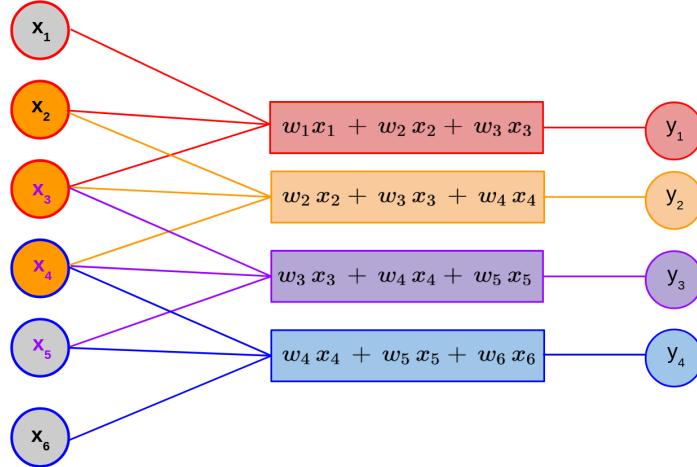


Figure 6.2: Representation of a 1D convolution layer in a neural network with a kernel size of 3 and a stride of 1. Here, no padding is used. The kernel (of size 3) slides over the neurons, summing and biasing the output of a given subset, extracting local information.

layer. One can perform pooling after every or several convolutions. Between convolution blocks, one can also use activation functions. Finally, the last layer of the network is fully connected. Examples of networks having this type of architecture are VGG16 [231], AlexNet [232] and YOLOv1 [233].

6.2.2 Making Deeper Networks: Residual Connections

While CNNs have been used in various contexts, they are not always optimal, especially for complex tasks when the networks become very large. In this case, one needs to make sure the addition of new layers leads to more expressivity and not only an increased computational complexity. If \mathcal{G} represents the ensemble of configurations the NN can reach, and we denote by g a given network configuration, then we want to find g^* the optimal one. Generally, this solution is not in \mathcal{G} and we look for $g_{\mathcal{G}}^*$, the best possible setting in \mathcal{G} . This is found by minimizing the loss. Now, we want to find \mathcal{G}' such that the optimal configuration is $g_{\mathcal{G}'}^*$ and is better than $g_{\mathcal{G}}^*$. This can happen only if $\mathcal{G} \subseteq \mathcal{G}'$. So, we need to construct a NN such that the addition of the new layer leads to a configuration encapsulating the previous one, preferably without significantly increasing the computational cost. For deep NNs, it is done by using *residual connections* (see Fig. 6.3 for a representation), where one applies an identity function on the layer's input in parallel to the usual operations [234]. Since the new network is a superset of the previous one, it can perform better. Additionally, the identity function is simple, not increasing the computational cost too much.

In practice, this is equivalent to learning the residuals of the mapping, which is done

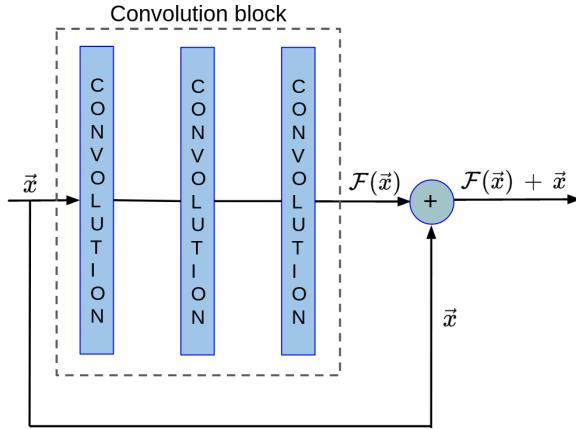


Figure 6.3: Representation of a residual connection. In parallel to the convolution block, we keep the input feature vector and sum it to the block's output. It is a way to learn the residuals of the transformation and build deeper networks without significantly increasing the computational cost.

through the residual connection [234]. The latter makes the network learn the mapping $\mathcal{F}(\vec{x})$, where \mathcal{F} represents the operation of a given block in the NN, such that

$$\vec{y} = \mathcal{F}(\vec{x}) + W\vec{x} \quad (6.13)$$

is the output of the convolution block and the residual connection. \mathcal{F} in Eq. (6.13) is a function of the weights and biases present in the layers, and W is an operation rescaling the input \vec{x} to have the same dimension as $\mathcal{F}(\vec{x})$. Since we do not do any other operations, the network does not become heavier, and the training efficiency is conserved. Learning the residuals from the mapping rather than the mapping decreases the training error efficiently while also improving the accuracy by permitting the design of larger NNs [234].

Such residual connections have been used to upgrade CNNs in image recognition tasks. For example, it is employed by YOLOv4 [235], an update of the YOLO algorithm with improved precision compared to the previous iteration with only convolution layers. Another architecture using residual connections is Inception-ResNETv4 [236]. In addition to residual connections, they use *inception blocks*, made of several branches where convolutions with different kernel sizes are performed in parallel. Therefore, a given block can look for correlations on different scales at once.

6.2.3 Batch Normalization

To reduce overfitting and improve the NN's performance, one can also introduce *batch normalization* [237]. It is a technique used to normalize the activation of the neurons in the network. One can see it as an additional layer made of two trainable parameters and modifying the output of a neuron to have a normalized mean and variance as input for the next layer. Schematically,

$$y = \gamma \frac{x - \mathbb{E}[x]}{\sqrt{\text{Var}[x] + \varepsilon}} + \beta, \quad (6.14)$$

where x is the output of the previous neuron, ε is a small constant added to avoid division by zero, γ and β are trainable parameters learned with the other network parameters, and y is the input for the next layer. This additional normalization enables the network to learn more efficiently [237].

6.3 Neural Networks for Probability Density Estimation

One of the main concerns in GW data analysis is the estimation of the binaries' parameters. This Bayesian inference problem would also benefit from the NN's speed after training as traditional methods (see Secs. 4.3 and 4.4) can be slow. For this, one can use *normalizing flows* (NFs). The main idea is to start from an initial simple probability distribution and pass it through a series of (simple) transformations to get the arbitrarily-complex probability distribution of interest. These transformations are parametrized with trainable parameters adapted through a training process to represent the distribution most faithfully – see Ref [238] for a comprehensive introduction.

6.3.1 Transforming Simple to Complex Probability Distributions

The main goal is to find a way to estimate the probability distribution $p(\mathbf{x})$ for a given multi-dimensional random variable \mathbf{x} . A NF represents \mathbf{x} as a transformation T of \mathbf{u} , where \mathbf{u} is sampled from a simple probability density $p(\mathbf{u})$:

$$\mathbf{x} = T(\mathbf{u}) \text{ and } \mathbf{u} \sim p(\mathbf{u}). \quad (6.15)$$

In this context, $p(\mathbf{u})$ is called the *base distribution*, and $p(\mathbf{x})$ the *target distribution*. The base distribution is generally an easy function parametrized by a set of parameters ϕ .

The transformations are parametrized by ψ . The ensemble of parameters $\theta = \{\phi, \psi\}$ represents trainable parameters, adjusted through a learning process¹.

The transformation T in Eq. (6.15) has to be invertible, leading to an inverse transformation T^{-1} . In the training process, one needs to go from \mathbf{x} to \mathbf{u} , while for the inference process, it is the opposite. In addition, T and T^{-1} have to be differentiable, meaning they are diffeomorphisms. Under these conditions, the probability distributions for the two variables are linked via the change of variable rule [238]

$$p(\mathbf{x}) = p(\mathbf{u}) |\det(J_T(\mathbf{u}))|^{-1}, \quad (6.16)$$

where $\mathbf{u} = T^{-1}(\mathbf{x})$, and $J_T(\mathbf{u})$ is the Jacobian of the transformation, corresponding to the matrix

$$J_T(\mathbf{u}) = \begin{pmatrix} \frac{\partial T_1}{\partial u_1} & \dots & \frac{\partial T_N}{\partial u_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial T_1}{\partial u_N} & \dots & \frac{\partial T_N}{\partial u_N} \end{pmatrix}, \quad (6.17)$$

where N is the dimensionality of \mathbf{x} and \mathbf{u} .

Equivalently, Eq. (6.16) can be written as

$$p(\mathbf{x}) = p(T^{-1}(\mathbf{x})) |\det(J_{T^{-1}}(\mathbf{x}))|. \quad (6.18)$$

Instead of using a single transformation and its inverse to get from one variable to the other, one often uses a series of simple transformations $\{T_k\}_{k=0,\dots,K}$ such that

$$T = T_0 \circ T_1 \circ \dots \circ T_K, \quad (6.19)$$

where T_k transform the variable \mathbf{z}_{k-1} into \mathbf{z}_k , assuming $\mathbf{z}_0 = \mathbf{u}$ and $\mathbf{z}_K = \mathbf{x}$. It relies on the composability of diffeomorphic transformations. This is where the term flow comes from: our variables flow through the transformations with a preserved normalized (probability) density like a fluid through a pipe.

The decomposition (6.19) can always be done for diffeomorphic transformations since they are always composable. Indeed, if we have two such transformations T_1 and T_2 , then $T_1 \circ T_2$ is also invertible and differentiable. In addition, its inverse is [238]

$$(T_1 \circ T_2)^{-1} = T_2^{-1} \circ T_1^{-1}, \quad (6.20)$$

and the determinant of its Jacobian is [238]

$$\det(J_{T_1 \circ T_2}(\mathbf{u})) = [\det(J_{T_1}(T_2(\mathbf{u})))][\det(J_{T_2}(T_1(\mathbf{u})))]. \quad (6.21)$$

Hence, everything can be computed based on individual diffeomorphic functions. Additionally, it is possible to show that for any variables $\mathbf{x} \sim p(\mathbf{x})$ and $\mathbf{u} \sim p(\mathbf{u})$, there exists

¹Here, we look at the most generic case. However, the base distribution can also be static, and only the transformation parameters ψ are adjusted during training.

a transformation T such that $\mathbf{x} = T(\mathbf{u})$, under reasonable conditions on $p(\mathbf{x})$ ² [238].

NFs are related to traditional ML approaches. Indeed, the different simple transformations have tunable parameters $\boldsymbol{\theta}$. Finding the optimal coefficients is the same as a fitting problem, where we want the final distribution $p(\mathbf{x}|\boldsymbol{\theta})$ to be a good representation of the theoretical distribution $p^*(\mathbf{x})$. Again, the model's parameters are tuned to have a minimal difference between the target distribution and the one provided by the NN.

For example, one can use the KL-divergence (see Sec. 6.1.2) between the two distributions to evaluate how good an approximation $p(\mathbf{x}|\boldsymbol{\theta})$ is. The loss function is then [238]

$$\mathcal{L}(\boldsymbol{\theta}) = D_{\text{KL}}[p^*(\mathbf{x})|p(\mathbf{x}|\boldsymbol{\theta})] = -\mathbb{E}_{p^*(\mathbf{x})} \left(\ln p(T^{-1}(\mathbf{u})|\boldsymbol{\psi}) + \ln |\det J_{T^{-1}}(\mathbf{x}|\boldsymbol{\theta})| \right) + C, \quad (6.22)$$

where C is a constant. If one has a set of samples $\{\mathbf{x}\}_{n=1,\dots,N} \sim p^*(\mathbf{x})$, then the loss function can be expressed by

$$\mathcal{L}(\boldsymbol{\theta}) \simeq -\frac{1}{N} \sum_{i=1}^N \ln p(T^{-1}(\mathbf{x}_i)|\boldsymbol{\psi}) + \ln (\det J_{T^{-1}}(\mathbf{x}_i|\boldsymbol{\theta})) + C. \quad (6.23)$$

Taking the gradient of Eq. (6.23) for the model parameters leads to

$$\nabla_{\boldsymbol{\phi}} \mathcal{L}(\boldsymbol{\theta}) \simeq -\frac{1}{N} \sum_{i=1}^N \nabla_{\boldsymbol{\phi}} \ln p(T^{-1}(\mathbf{x}_i|\boldsymbol{\psi})) + \nabla_{\boldsymbol{\phi}} \ln |\det J_{T^{-1}}(\mathbf{x}_i|\boldsymbol{\phi})|, \quad (6.24)$$

$$\nabla_{\boldsymbol{\psi}} \mathcal{L}(\boldsymbol{\theta}) \simeq -\frac{1}{N} \sum_{i=1}^N \nabla_{\boldsymbol{\psi}} \ln p(T^{-1}(\mathbf{x}_i)|\boldsymbol{\psi}). \quad (6.25)$$

So, only the inverse transformation and its associated Jacobian are needed to train the model, provided we have access to samples from the target distribution. Minimizing the KL divergence is equivalent to maximizing the log-likelihood of the target distribution under the transformation. However, one cannot always have samples from the target distribution beforehand. Therefore, it is convenient to rewrite the loss function (6.22) as

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{p(\mathbf{u}|\boldsymbol{\psi})} \left[\ln p(\mathbf{u}|\boldsymbol{\psi}) - \ln |\det J_T(\mathbf{u}|\boldsymbol{\phi})| - \ln p^*(T(\mathbf{u})|\boldsymbol{\theta}) \right], \quad (6.26)$$

of which one can take derivatives with respect to the model parameters to perform gradient descent. Here, one can evaluate the expression as long as the target distribution is evaluable for given samples. One also needs $p(\mathbf{u})$, T and its determinant.

In Eq. (6.26), it is enough to know the true distribution up to a constant. So, if we know $\tilde{p}(\mathbf{x})$ such that $p^*(\mathbf{x}) = \tilde{p}(\mathbf{x})/C$, with C some constant, we can compute the

²For example a non-zero probability on the considered domain for \mathbf{x} .

loss via Eq. (6.26). This is easier because, according to Bayes' theorem (4.14), the posterior probability can be expressed as a product of the likelihood and the priors for the sample, divided by the evidence. The latter is nothing else than a normalization constant. Therefore, the equation can be evaluated if we have a modeled likelihood and some known priors. Finally, if one can sample $p(\mathbf{u})$ directly, the second loss function makes it possible to train the model without having to know the exact shape of the posteriors in advance. Therefore, we can train a NN to perform Bayesian analysis.

There are various possibilities to model the different transformation mapping $p(\mathbf{u})$ to $p(\mathbf{x})$ [238]. For example,

- *Autoregressive flows* [239]: made by two main elements, the transformer, and the conditioner. The first is transforming the samples from one step to the other, while the second controls how the transformation happens. One has then some freedom in choosing the exact characteristics of the two elements (see [238] for examples). These flows are usually expressive but have the disadvantage of being slow in the forward or backward direction [239].
- *Linear flows*: alleviates one of the issues of autoregressive flows, where the order of the input variables generally matters. Here, the transformation from one step to the other is parametrized by an invertible matrix. However, it is less expressive since one only has linear transformations [238].
- *Coupling flows* [240, 241]: make the forward and backward operations fast but reduce a bit the expressivity of the individual layers [238]. The idea is to split the network parameters into two subsets arbitrarily. The first subset is unchanged and the layer just outputs the initial value, while the second subset is transformed using an invertible function. It can be shown that stacking multiple coupling layers enables one to get a network able to represent any distribution [238].

The main idea behind these structures is to construct a tractable Jacobian, done by representing it as a triangular matrix. Autoregressive and coupling flows are two ways of obtaining this, making them common flow algorithms.

6.3.2 Continuous Normalizing Flows

In the previous section, the transformations are cast as a series of discrete steps taken one after the other. An alternative, and often more efficient approach, is to use infinitesimal steps to make the entire transformation a result of the integration of all the smaller steps [242]. This is called a *continuous NF* (CNF). Here, the flow's evolution is described by ordinary differential equations (ODEs). By analogy with fluids, one can see the parameters describing the model flowing through the different steps. Therefore, one can

parametrize the ODEs using time and see the network as making the samples evolve to map $p(\mathbf{u})$ to $p(\mathbf{x})$.

If the state of the flow at time t is written \mathbf{z}_t , with $\mathbf{z}_{t_0} = \mathbf{u}$ and $\mathbf{z}_{t_1} = \mathbf{x}$, and t evolves continuously, the transformation is described by

$$\frac{d\mathbf{z}_t}{dt} = g(t, \mathbf{z}_t | \boldsymbol{\theta}), \quad (6.27)$$

where g is a function representing the transformation and parametrized by $\boldsymbol{\theta}$, corresponding to the NN's parameters. Eq. (6.27) has a unique solution if g is Lipschitz continuous³ in \mathbf{z}_t and continuous in t [242]. Many NN layers respect these criteria, making it relatively straightforward to build a CNF.

The full transformation T such that $\mathbf{x} = T(\mathbf{u})$ is given by

$$\mathbf{x} = \mathbf{u} + \int_{t_0}^{t_1} dt g(t, \mathbf{z}_t | \boldsymbol{\theta}), \quad (6.28)$$

and the inverse transformation by

$$\mathbf{u} = \mathbf{x} + \int_{t_1}^{t_0} dt g(t, \mathbf{z}_t | \boldsymbol{\theta}). \quad (6.29)$$

Therefore, a CNF has the same computational complexity in both directions. It is not necessarily the case for discrete normalizing flows.

Working in log space, the evolution of the log probability density is [242]

$$\frac{d \ln p(\mathbf{z}_t)}{dt} = -\text{Tr}[J_g(\mathbf{z}_t)], \quad (6.30)$$

where J_g is the Jacobian of the transformation using g . For high dimensions, the trace can be approximated as [243–245]

$$\text{Tr}[J_g(\mathbf{z}_t)] \simeq \mathbb{E}[\mathbf{v}^T J_g(\mathbf{z}_t) \mathbf{v}], \quad (6.31)$$

where \mathbf{v} is an N -dimensional (N being the number of dimensions in the problem) random vector with $\mathbb{E}[\mathbf{v}] = 0$ and $\mathbb{E}[\mathbf{v}^T \mathbf{v}] = 1$. While the computational complexity of the trace is already lower than for the determinant, this approximation reduces it even more. So CNFs recast the problem in a less computationally expensive way.

Using Eq. (6.30), the final probability density is obtained via an integration similar to the one in Eq. (6.28), giving

$$\ln p(\mathbf{x}) = \ln p(\mathbf{u}) - \int_{t_0}^{t_1} dt \text{Tr}[J_g(\mathbf{z}_t)]. \quad (6.32)$$

³A function is Lipschitz continuous if there exists a real R such that for any pair of points in the function's domain, the slope of the line connecting the two points is smaller than R .

In the end, one can compute the variable \mathbf{x} and its (log) probability simultaneously by considering the system of equations made of Eqs. (6.28) and (6.32). The system has to be solved numerically. Efficient numerical integration methods exist and can be used. For example, there are the well-known Euler and the Adjoint methods [238]. Other, faster and more robust methods have also been developed, for example, the asynchronous leapfrog (ALF) method [246] or the memory efficient ALF integrator (MALI) [247].

The rest of the procedure is similar to the discrete case, and one can directly adjust the weights by minimizing the KL divergence.

6.3.3 Accounting for Data in Normalizing Flows

Secs. 6.3.1 and 6.3.2 show how to convert a simple distribution into a more complex one via transformations modeled with a NN. However, when solving a Bayesian inference problem, one often wants to account for the data. Therefore, one introduces *conditional* NFs [248], where one makes the different distributions conditional on the data. So, Eq. (6.16) becomes

$$p(\mathbf{x}|d) = p(\mathbf{u}|d) \det |J_T(\mathbf{u}, d)|^{-1}, \quad (6.33)$$

where d is the data. In this expression, everything is conditioned on the data, even the prior. The latter is not necessary, and one can use a static prior.

The conditional NFs and CNFs can be constructed the same way as before, where the transformation T accounts for the conditionality on the data. Therefore, the final probability density probed also becomes a function of the data and can be evaluated using the data-dependent likelihood and priors on \mathbf{x} . The loss function is still Eq. (6.26), now with probabilities conditional on the data, and the training can be done as before. The main difference is that one needs to have many data sets to cover the full possible parameter space and learn to model the corresponding probability distributions.

Putting everything together, it is possible to construct a conditional continuous NF network, mapping parameters \mathbf{u} to the posteriors $p(\mathbf{x}|d)$ using some transformation function $g(\mathbf{z}_t, t|\boldsymbol{\theta}, d)$. The latter describes the evolution of the samples with time, is parametrized by the NN parameters $\boldsymbol{\theta}$, and depends on the input data d . This approach enables one to benefit from the efficiency of CNFs while also making the posterior data dependent. Such a framework fulfills the requirements to perform Bayesian inference.

6.4 Pre- and Post-Processing the Data

A NN starts with an input layer taking in the data. One can decide to use the raw data. However, this is not always the best option. First, the data can be very long,

and one would need a very high number of input neurons. Additionally, the first layers of the network are used to interpret the data instead of analyzing it, losing efficiency. Therefore, it can be better to modify the data first and give the network a reduced version of it. In addition, it can happen that the network's performance is not optimal. In that case, one can apply a post-processing step to correct and improve the output. If the post-processing method is well chosen, it can also alleviate the black-box nature of ML. In this section, we review some methods used to summarize the data before passing it to the network as well as post-processing steps to improve the network's output.

6.4.1 Normalizing the Data

GW data has quite often significant relative amplitude variations, with signals having an amplitude much smaller than the noise they are buried in. Passing this raw data directly to the NN is usually not a suitable option as the numerous features can lead to confusion. Therefore, one can apply some pre-processing steps. A first such step is *whitening*, where one performs a normalization of the data with the noise estimate, dividing the frequency-domain strain by the square root of the PSD. It is a good way to remove some noise artifacts from the data. For example, there are known peaks in the PSD (see Fig. 4.1 for an illustration) which have large contributions in the frequency power but are due to noise.

Additionally, NNs generally prefer having a normalized input. A possible way to do this is to normalize the data between -1 and 1 by dividing by the absolute maximum. However, if there is one dominant value in the data, it can lead to tiny values for most other data points. Instead, one can *standardize* the data, where the goal is to transform the data so that it has a zero mean and a unit variance. For example, one can do this by fitting a Gaussian through the data. This exact method is data-specific and can be applied batch-wise.

6.4.2 Extracting Essential Features of the Data

In some cases, standardizing the input is not enough to have a well-performing NN. Noise and spurious features can still lead to confusion. Additionally, the data can remain relatively heavy, and the number of sample points considered is unchanged. Therefore, it is sometimes more advantageous to extract summary information from the data and use this as input for the NN. While recent developments have shown that this extraction could be optimized using NNs [249], the method explained here, and used in this thesis, is a classical one, called *single value decomposition* (SVD) [250, 251]. The idea is to reduce the number of elements representing the data by finding a basis of orthogonal

vectors on which to decompose the signals. Additionally, this has the advantage of reducing the noise passed to the network, enabling it to focus more on the signal.

If \mathbf{d} is the array of detector strain data, we want to find basis vectors \mathbf{u}_i and a reconstruction matrix \mathbf{A} such that the SNR for a template h_j is

$$\rho_j = \sum_{i=1}^N A_{ji} (\mathbf{u}_i \cdot \mathbf{d}). \quad (6.34)$$

If the data under consideration is an array of waveforms $\{h_1, h_2, \dots, h_M\}$, we seek an SVD matrix \mathbf{H} such that [251]

$$H_{ij} = \{\Re h_1, \Im h_1, \Re h_2, \Im h_2, \dots, \Re h_M, \Im h_N\}, \quad (6.35)$$

where row vectors \vec{H}_{2i-1} and \vec{H}_{2i} represent the real and the imaginary parts of the i^{th} template in the set.

The SVD decomposition for the matrix (6.35) is [251]

$$H_{ij} = \sum_{k=1}^N u_{ik} \sigma_k v_{kj}, \quad (6.36)$$

where \mathbf{v} is an orthonormal matrix containing the reconstruction coefficients such that

$$\sum_i v_{ij} v_{ik} = \delta_{jk}. \quad (6.37)$$

σ is a vector containing the singular values representing the data and ordered by importance in the template reconstruction. \mathbf{u} is a matrix representing vectors of an orthonormal basis with

$$\sum_i u_{ij} u_{ik} = \delta_{jk}. \quad (6.38)$$

CBC search sensitivities are such that an error of a few percent is undetectable. Therefore the matrix H_{ij} can be represented with a truncated decomposition, using $N' < N$ elements for its representation

$$H_{ij} \simeq \sum_{k=1}^{N'} v_{ik} \sigma_k u_{kj}. \quad (6.39)$$

One can then evaluate the reconstructed SNR for the signal as

$$\rho'_j = \sum_{k=1}^{N'} [v_{(2j-1)k} \sigma_k i v_{(2j)k}] (\mathbf{u}_k \cdot \mathbf{d}). \quad (6.40)$$

One can truncate the decomposition at a given precision level, corresponding to selecting the X first elements of σ . In addition, it is often enough to use the few hundred first

elements, significantly reducing the amount of data passed to the NN, and the elements summarize key features of the signal. Therefore, they help the NN focus on important characteristics, making the learning process lighter and more efficient. An example of an algorithm used to perform SVD decomposition can be found in Ref. [250]. Performing the SVD decomposition on noisy data reduces the noise and makes the NN more sensitive to the GW signal present in the data.

6.4.3 Linking Machine Learning With Classical Methods

One concern often faced when using ML is its black-box nature. It is possible to reduce this concern by comparing the final samples obtained using ML with those obtained using MCMC or nested sampling. This can be done via *importance sampling*.

If q is a probability distribution covering the same space as another distribution p , we can transform q into p provided that (i) p is known up to a normalization constant, hence we know \tilde{p} such that $p(\boldsymbol{\theta}) = \tilde{p}(\boldsymbol{\theta})/Z$, and (ii) $q(\boldsymbol{\theta}) \neq 0$ for all $\boldsymbol{\theta}$ such that $p(\boldsymbol{\theta}) \neq 0$. Then, we can compute the ratio

$$w(\boldsymbol{\theta}) = \frac{\tilde{p}(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}, \quad (6.41)$$

which enables us to transform the base distribution into the target distribution. One can see $w(\boldsymbol{\theta})$ as a measure of the (di)similarity between the two distributions under consideration. $w(\boldsymbol{\theta})q(\boldsymbol{\theta})$ eventually converges to $p(\boldsymbol{\theta})$ when enough samples are used [252].

For GW data analysis, $\tilde{p}(\boldsymbol{\theta})$ can be obtained via the likelihood as in Eq. (4.18), and $q(\boldsymbol{\theta})$ are the samples provided by ML. As long as the description given by ML is broad enough to cover the traditional methods' posterior space, importance sampling can be used to transform one into the other. This alleviates the black-box issue as we have a measure of the difference between ML and Bayesian techniques.

Part II

GRAVITATIONAL-WAVE LENSING:
MAGNIFYING GENERAL RELATIVITY

7

SEARCHES FOR STRONG GRAVITATIONAL-WAVE LENSING AND RELATED ISSUES

Strong lensing and microlensing were first searched for by the LIGO-Virgo-KAGRA (LVK) collaboration in 2019 [172], but this pioneering paper reported no detection. At the time, the strong-lensing searches were based only on the posterior overlap technique [181], introduced in Sec. 5.3. So, they only looked for overlapping posteriors in a sub-group of parameters. This can lead to a significant false alarm risk, but it is a good first indicator. On the strong lensing side, the GW170104-GW170814 event pair is the most interesting candidate. However, even if it has consistent posterior distributions, it has a very long time delay, incompatible with galaxy-lensing, which is the most likely type of lensing to be observed [172].

Complementing the first searches, other more precise methods were developed. They also use the lensing parameters linking the two events and analyze the two data streams jointly under the lensed hypothesis [214, 215]. In Ref. [215], the authors re-analyze the O2 data, confirming GW170104-GW170814 as an interesting pair with a high significance. Still, the evidence is too low to state the event pair is strongly lensed. The interest in this pair is also followed up by groups outside of the LVK Collaboration, as attested by [189]. In this study, the authors also flag this event but without enough evidence to claim lensing. They find an additional sub-threshold event compatible with the two first events, making it a possible triplet. However, even with the addition of a third event, we do not have enough evidence to claim these are strongly-lensed images.

In parallel, the GW lensing field has continued evolving, developing interesting applications. For example, one can note the emerging interest in multi-messenger strong lensing, where one would link the observed GW data with an EM counterpart [192, 193]. From such an observation, one could probe cosmology [192] or test fundamental physics [198, 199]. At the same time, the increased number of detectors in the fiducial detector network obtained through the observation of multiple images opens the possibility of probing the GW polarization more efficiently [200].

During the first half of the third LVK observation run, called O3a, lensing searches have continued without any significant detection [188]. Thanks to the development of new search methods, more complete analyses were performed on the data. On the strong lensing side, one now searches for strong lensing in several ways. First, one looks at unusually high-mass systems. Since the luminosity distance is biased by the

magnification, the inferred redshift is also biased. As a consequence, the inferred source-frame component masses are biased. Therefore, we would seemingly observe events with masses higher than those expected according to population models (see, for example, [253] for BNS systems). Searches for multiple images due to strong lensing are now multi-staged, with a first search done by posterior overlap to flag candidates rapidly. Interesting pairs are followed up by two different joint parameter estimation pipelines: one is LALINFERENCE-based [215] and the other is BILBY-based [214]. The second pipeline also includes selection effects and is, therefore, used to assess the lensed nature of the events. Additionally, one looks for sub-threshold counterparts of the detected events; one makes a sub-bank of templates based on the observed super-threshold events' characteristics and sees if any events with an $\text{SNR} < 8$ are detected [254, 255]. Microlensing searches were also performed on the O3a data. None of these searches led to a detection [188], enabling us to place constraints on the lensing rate, the merger rate, and the dark-matter content of the Universe [188].

All these efforts are motivated by the predicted rate for strong lensing, with a non-negligible rate in the coming years [180, 187]. However, all methods to look for strong lensing [181, 214, 215] rely on analyzing pairs of events. This can lead to issues since the number of detected events rapidly grows with detector upgrades, reaching a predicted $\mathcal{O}(1000)$ events in O5 [180], leading to $\mathcal{O}(5 \times 10^5)$ pairs to analyze. This number will only increase in the coming years when the current detectors get upgraded and new ones join the ground-based interferometer network. Therefore, the search methods have to be fast.

Other works have also shown that major false alarm risks are linked to the increasing number of detected events [187, 201]. Indeed, as the number of events grows, the chance of having matching binary parameters by chance increases, leading to seemingly lensed events. One way to decrease this risk is to make the search pipelines more accurate, using joint parameter estimation, for example. However, these methods are too slow to follow the number of pairs that need to be analyzed. Another approach is to account for the expected values of the lensing parameters by including so-called lensing statistics [181, 187, 203]. This helps discard more events that overlap by chance by identifying those with apparent lensing parameters incompatible with the expected values. However, this approach is model-dependent, and if one is not considering the correct lens model, it can lead to the non-identification of genuinely lensed events.

In this part of the thesis, we present work done to address some of the issues related to the identification and characterization of strongly-lensed GW events.

8

GOLUM: A FAST AND PRECISE METHODOLOGY TO SEARCH FOR AND ANALYZE STRONGLY LENSED GRAVITATIONAL WAVES

In Ref. [191], we developed a method to search for strongly-lensed GWs combining speed and precision, theoretically equivalent to joint parameter estimation under the lensed hypothesis. It is based on an alternative way to write the coherence ratio (5.32), making the evidence for one image conditioned on the other image. This leads to a “*distribution*” of the parameter estimation runs. In this chapter, we explain in detail how this works. Combining speed and precision should enable one to analyze more signals and keep up with the forecast detection rates without increasing the related false-alarm probability (FAP) as much as with posterior overlap.

8.1 Recasting the Lensed Evidence for Fast Computation

For strong lensing, the amplification factor is given by Eq. (5.15). Therefore, the lensed and unlensed waveforms are related as [169]

$$\tilde{h}_L(\boldsymbol{\theta}, \mu_j, t_j, n_j) = \sqrt{\mu_j} e^{(2i\pi f t_j - i\pi n_j \text{sign}(f))} \tilde{h}_U(\boldsymbol{\theta}), \quad (8.1)$$

where $\boldsymbol{\theta}$ are the usual BBH parameters, μ_j is the relative magnification, t_j the time delay, and n_j the Morse factor. These three additional parameters represent the lensing parameters for image j . However, the magnification and the time delay are degenerate with the luminosity distance and the coalescence time, respectively. Therefore, one would observe an *apparent* luminosity distance and an *apparent* time delay

$$D_L^{\text{app}} = \frac{D_L}{\sqrt{\mu_j}} \quad (8.2)$$

$$t_c^{\text{app}} = t_c + t_j, \quad (8.3)$$

where D_L and t_c are the unlensed luminosity distance and coalescence time, respectively.

Because of these degeneracies, one uses relative lensing parameters, linking the observed values between the images. Assuming a pair of events, one has

$$\begin{aligned} t_c^{\text{app},2} &= t_c^{\text{app},1} + t_{21}, \\ D_L^{\text{app},2} &= \sqrt{\mu_{\text{rel}}} D_L^{\text{app},1}, \\ n_2 &= n_1 + n_{21}, \end{aligned} \quad (8.4)$$

where t_{21} is the time delay between the two images, μ_{rel} is the relative magnification, i.e. the ratio of the individual magnifications, and n_{21} is the Morse factor difference. These parameters link the lensed waveforms for the two images.

To find whether two events are lensed or not, one can compute the coherence ratio (5.32), which is the ratio of the lensed (5.34) and unlensed (5.33) evidence. The lensed evidence described by Eq. (5.34) requires jointly analyzing the two data streams, which can be cumbersome [214, 215]. However, the lensed evidence can be recast into a *conditioned* evidence. As explained here, this new expression decreases the computational cost related to joint parameter estimation.

The joint evidence under the lensed hypothesis can be rewritten as

$$p(d_1, d_2 | \mathcal{H}_L) = p(d_2 | d_1, \mathcal{H}_L) p(d_1 | \mathcal{H}_L). \quad (8.5)$$

We have

$$p(d_1 | \mathcal{H}_L) = \int \int d\mathbf{n}_1 d\boldsymbol{\Theta} p(d_1 | \boldsymbol{\Theta}, n_1, \mathcal{H}_L) p(n_1, \boldsymbol{\Theta} | \mathcal{H}_L), \quad (8.6)$$

with n_1 the Morse factor for the first image and $\boldsymbol{\Theta}$ the usual BBH parameters with the luminosity distance and coalescence time replaced by the apparent ones for the first image.

The joint evidence can then be expressed as

$$p(d_2 | d_1, \mathcal{H}_L) = \int \int d\boldsymbol{\Theta} d\boldsymbol{\phi} p(d_2 | \boldsymbol{\Theta}, \boldsymbol{\phi}, \mathcal{H}_L) p(\boldsymbol{\Theta} | d_1, \mathcal{H}_L) p(\boldsymbol{\phi} | \mathcal{H}_L), \quad (8.7)$$

where $\boldsymbol{\phi} = \{\mu_{\text{rel}}, t_{21}, n_{21}\}$ is the array of relative lensing parameters.

Eq. (8.5) is recovered through the repeated use of Bayes' theorem:

$$\begin{aligned} p(d_1, d_2 | \mathcal{H}_L) &= \int \int d\boldsymbol{\Theta} d\boldsymbol{\phi} p(d_1, d_2 | \boldsymbol{\Theta}, \boldsymbol{\phi}, \mathcal{H}_L) p(\boldsymbol{\Theta} | \mathcal{H}_L) p(\boldsymbol{\phi} | \mathcal{H}_L) \\ &= \int \int d\boldsymbol{\Theta} d\boldsymbol{\phi} p(d_2 | d_1, \boldsymbol{\Theta}, \boldsymbol{\phi}, \mathcal{H}_L) p(d_1 | \boldsymbol{\Theta}, \mathcal{H}_L) p(\boldsymbol{\Theta} | \mathcal{H}_L) p(\boldsymbol{\phi} | \mathcal{H}_L) \\ &= p(d_1 | \mathcal{H}_L) \int \int d\boldsymbol{\Theta} d\boldsymbol{\phi} p(d_2 | d_1, \boldsymbol{\Theta}, \boldsymbol{\phi}, \mathcal{H}_L) p(\boldsymbol{\Theta} | d_1, \mathcal{H}_L) p(\boldsymbol{\phi} | \mathcal{H}_L) \\ &= p(d_1 | \mathcal{H}_L) \int \int d\boldsymbol{\Theta} d\boldsymbol{\phi} p(d_2, \boldsymbol{\Theta}, \boldsymbol{\phi} | d_1, \mathcal{H}_L) \\ &= p(d_1 | \mathcal{H}_L) p(d_2 | d_1, \mathcal{H}_L). \end{aligned} \quad (8.8)$$

Eqs. (8.6) and (8.7) can be used to explicitly write the coherence ratio

$$\mathcal{C}_U^L = \frac{p(d_1|\mathcal{H}_L) p(d_2|d_1, \mathcal{H}_L)}{p(d_1|\mathcal{H}_U) p(d_2|\mathcal{H}_U)} \quad (8.9)$$

in terms of likelihoods. The labels 1 and 2 given to the data are arbitrary and can be swapped if desired. A good practice could be to use the best-observed image (highest SNR, most detectors active at the detection time, ...) as the first image as it is the one that best constrains the event's parameters.

The analysis for the first image, computing Eq. (8.6) is a usual BBH run with an additional Morse factor included. For the second image, the equation's dimensionality has been reduced from 16 to 3 dimensions since only the unlensed parameters are sampled from the first image's posteriors. Additionally, since the BBH parameters are already known via the first image run, these parameters are concentrated in the correct region of space. Therefore, the second image run is faster.

8.2 Evaluating the Conditioned Evidence

In principle, one can develop several ways to solve the conditioned evidence. Our approach is to rewrite the conditioned evidence in terms of a “*marginalized likelihood*” as

$$p(d_2|d_1, \mathcal{H}_L) = \int d\phi L(\phi|\mathcal{H}_L) p(\phi|\mathcal{H}_L), \quad (8.10)$$

where

$$L(\phi|\mathcal{H}_L) = \left\langle p(d_2|\Theta, \phi, \mathcal{H}_L) \right\rangle_{p(\Theta|d_1, \mathcal{H}_L)}. \quad (8.11)$$

It is the likelihood of the second image averaged over the posterior samples of the first one. Evaluating this expression is desirable for the speed of the analysis since the posteriors for Θ are already focused on the correct region of the parameter space.

The passage from the conditioned evidence (8.7) to the marginalized likelihood (8.10) is performed as

$$\begin{aligned} p(d_2|d_1, \mathcal{H}_L) &= \int \int d\Theta d\phi p(d_2, \Theta, \phi|d_1, \mathcal{H}_L) \\ &= \int d\phi p(\phi|\mathcal{H}_L) \int d\Theta p(d_2|\Theta, \phi, \mathcal{H}_L) p(\Theta|d_1, \mathcal{H}_L) \\ &= \int d\phi \left\langle p(d_2|\Theta, \phi, \mathcal{H}_L) \right\rangle_{p(\Theta|d_1, \mathcal{H}_L)} p(\phi|\mathcal{H}_L) \\ &= \int d\phi L(\phi|\mathcal{H}_L) p(\phi|\mathcal{H}_L). \end{aligned} \quad (8.12)$$

Evaluating the condition evidence gives access to the posterior for the lensing parameters $p(\phi|d_1, d_2)$. Still, we only have the posteriors for Θ as a function of the first image. However, the observation of the second image should be impacted on these posteriors since it brings a major improvement in some parameters, such as the sky location for example [192]. This can be done through a reweighing process, using

$$p(\Theta, \phi|d_1, d_2) \propto \frac{p(d_2|\Theta, \phi)}{p(d_1, d_2|\phi)} p(\Theta|d_1) p(\phi|d_1, d_2). \quad (8.13)$$

Indeed, the probability to have both the source and the lensing parameters can be decomposed as

$$p(\Theta, \phi|d_1, d_2) = p(\Theta|\phi, d_1, d_2) p(\phi|d_1, d_2). \quad (8.14)$$

Further writing out the probability of having the (apparent) source parameters given the lensed ones and the two data streams gives

$$\begin{aligned} p(\Theta|\phi, d_1, d_2) &= \frac{p(d_1, d_2|\Theta, \phi) p(\Theta|\phi)}{p(d_1, d_2|\phi)} \\ &= \frac{p(d_2|d_1, \Theta, \phi) p(d_1|\Theta, \phi) p(\Theta|\phi)}{p(d_1, d_2|\phi)} \\ &\propto \frac{p(d_2|\Theta, \phi) p(\Theta|d_1)}{p(d_1, d_2|\phi)}. \end{aligned} \quad (8.15)$$

Finally, combining Eqs. (8.14) and (8.15) gives Eq. (8.13).

Eqs. (8.5) and (8.10) are exact expressions. However, to decrease the computational time, one often takes sub-samples from the first image posteriors before evaluating Eq. (8.10), leading to some approximation. If enough samples are selected, the method remains sufficiently precise.

8.3 Additional Likelihood Evaluation Speed up

The evaluation of Eq. (8.10) can further be sped up by using a look-up table. For this, one recycles the parameters from the first image and uses them to pre-compute the waveforms. Therefore, during the sampling process, one does not need to generate the full waveforms anymore as it suffices to correct for the effect of the relative lensing parameters.

Indeed, if $\{d_1^j\}$ and $\{d_2^j\}$ are the data sets corresponding to images 1 and 2, respectively, with j denoting the detector index, the log-likelihood of the second image can be expressed

as

$$\begin{aligned} 2 \ln L(\Theta, \phi) &= \sum_j \langle d_2^j - h_2^j(\Theta, \phi) | d_2^j - h_2^j(\Theta, \phi) \rangle \\ &= \sum_j \left(\langle d_2^j | d_2^j \rangle + \frac{1}{\mu_{\text{rel}}} \langle h_1^j(\Theta) | h_1^j(\Theta) \rangle - \frac{2}{\sqrt{\mu_{\text{rel}}}} \langle d_2^j | h_1^j(\Theta) \rangle e^{(2i\pi f t_{21} - i\pi n_{21} \text{sign}(f))} \right), \end{aligned} \quad (8.16)$$

where $\langle \cdot | \cdot \rangle$ is the inner product (4.7). The different inner products in this expression are a function of the first image's waveform only. Therefore, it can be computed as soon as samples are selected. During the second image run, for each set of samples ϕ , one can directly correct the pre-computed inner products with the lensing parameters, which makes the inference process significantly faster. The application of a look-up table reduces the computational time by a factor ~ 20 .

The new methodology formed by the conditional likelihood scheme and the look-up table is implemented in GOLUM (*Gravitational-wave analysis Of Lensed and Unlensed waveform Models*) [191, 256], which is a module of BILBY [149]. Since we also incorporate a lensed waveform, applying the Morse phase as an overall phase-shift, our framework works with any waveform model, ranging from an aligned-spin waveform model like *IMRPhenomD* [104], to a precessing-spin waveform model, like *IMRPhenomPv2*, and even to a precessing waveform with HOMs, like *IMRPhenomXPHM* [121]. With this framework, the second image run takes less than $\mathcal{O}(1)$ CPU-hour¹ using the PYMULTINEST sampler [153].

8.4 Example Analysis

As a practical example, we inject the signal from a spinning, precessing BBH merger generated with *IMRPhenomPv2*, with parameters listed in the second column of Table 8.1, into synthetic stationary, Gaussian noise for a network of 3 detectors (LIGO-Livingston, LIGO-Hanford, and Virgo) at design sensitivity [25, 26]. This event has a network SNR of ~ 23 . We then inject the event's lensed counterpart image, with relative magnification between the two events $\mu_{\text{rel}} = 2$, relative Morse factor $n_{21} = 0.5$, and relative time delay $t_{21} = 14$ hr. Throughout this analysis, we use a uniform prior for the relative magnification ($\mu_{\text{rel}} \in [0.01, 20]$), the time-delay ($t_{21} \in [t_{21} - 0.1, t_{21} + 0.1]$ s), the chirp mass (2.50) ($\mathcal{M}_c \in [10, 100]$ M $_{\odot}$), and mass ratio ($q \in [0.1, 1]$, with $q = m_2/m_1$); the spin distribution is isotropic. The prior on the Morse factor is a discrete uniform distribution over the three possible values ($n_1 \in \{0, 0.5, 1\}$), and the prior in Morse factor difference is a discrete uniform distribution over the four possible values ($n_{21} \in$

¹This is the total time to compute the evidence and joint posteriors with an *Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz* processor.

Parameter	Lensed event	Unlensed event
Mass 1 (m_1)	$36.0 M_{\odot}$	$35.8 M_{\odot}$
Mass 2 (m_2)	$29.2 M_{\odot}$	$11.4 M_{\odot}$
Spin amplitude 1 (a_1)	0.4	0.3
Spin amplitude 2 (a_2)	0.3	0.2
Tilt angle 1 (θ_1)	0.5	1.0
Tilt angle 2 (θ_2)	1.0	2.2
Spin vector azimuthal angle (ϕ_{12})	1.7	5.1
Precession angle about angular momentum (ϕ_{jl})	0.3	2.5
Luminosity distance (d_L) ^a	1500 Mpc	500 Mpc
Inclination angle (ι)	0.4	1.9
Wave polarization (ψ)	2.659	2.7
Unlensed phase of coalescence (ϕ_c)	1.3	3.7
Morse factor (n_1)	0.5	0
Right ascension (α)	1.375	3.9
Declination (δ)	-1.2108	0.22
Time of coalescence (t_c)	1126259642.413	10.04

Table 8.1: Summary of the injection parameters used for the example runs in Sec. 8.4. In this table and throughout the work, the angles are measured in radians.

^aIn the lensing framework, the distance of the event is the *apparent* one, as both images are, in fact, affected by the lensing parameters.

$\{0, 0.5, 1, 1.5\}$)². These priors do not include results from lens modeling of the inferred astrophysical BBH population.

To analyze this lensed pair, we perform four nested sampling runs. Firstly, we analyze the two injections under the non-lensed hypothesis. Secondly, we estimate the parameters of one of the events under the lensing hypothesis. Thirdly, we obtain the conditioned evidence $p(d_2|d_1, \mathcal{H}_L)$, by sampling the second event's likelihood based on the earlier lensed parameter estimation, thereby also obtaining the relative image properties (see Fig. 8.1).

Combining the four runs, we obtain the coherence ratio \mathcal{C}_U^L . In our example lensed simulation, we find $\log \mathcal{C}_U^L = 23.6$, correctly consistent with lensing.

We then inject two unrelated events (see the last column of Table 8.1 for the parameters) and repeat the analysis. In this case, the coherence ratio $\log \mathcal{C}_U^L = -14$, not consistent with lensing, as expected.

We can also combine information from the two lensed images to better constrain the binary parameters. In particular, we can use the posterior of the lensed parameters obtained from the combined run to reweigh the posterior samples of the first run as in

²The negative value $n_{21} = -0.5$ is equivalent to the transformation $n_{21} = 1.5$, and the negative value $n_{21} = -1$ is equivalent to $n_{21} = 1$. Thus, we do not consider them.

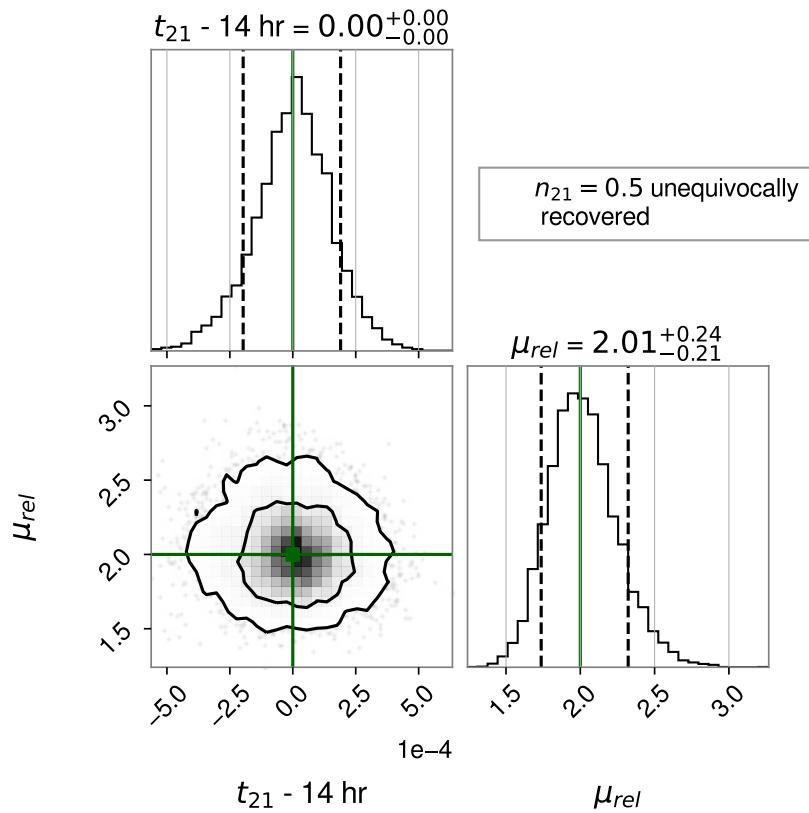


Figure 8.1: Posterior distribution of the magnification and time delay (re-centered at zero) between two strongly lensed GW images. The parameters are well recovered, and the difference in Morse factors is fully determined, allowing us to make inferences about the image properties. If the event pair is part of a strongly lensed quadruplet lensed by a galaxy, the first image type would likely be type-I and the second one type-II. The injected values are $\mu_{\text{rel}} = 2$, $n_{21} = 0.5$, and $t_{21} = 50400 \text{ s} = 14 \text{ hr}$.

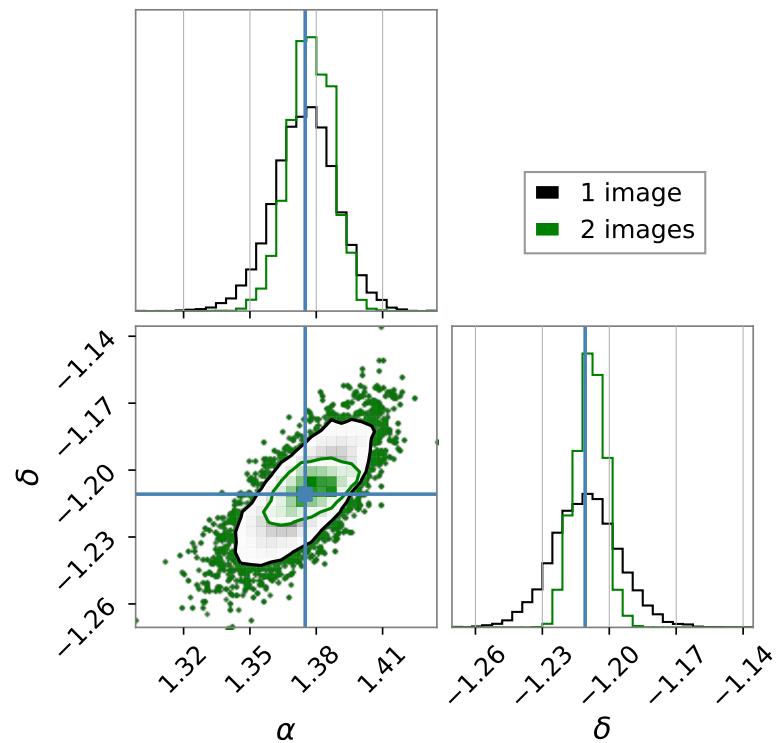


Figure 8.2: The combined sky location (right ascension α and declination δ) of two strongly lensed GW images. The black distributions refer to the posteriors from the analysis of the first image only, and the green distributions are the results when the two images are combined. The detection of several images significantly reduces the 90% credible region.

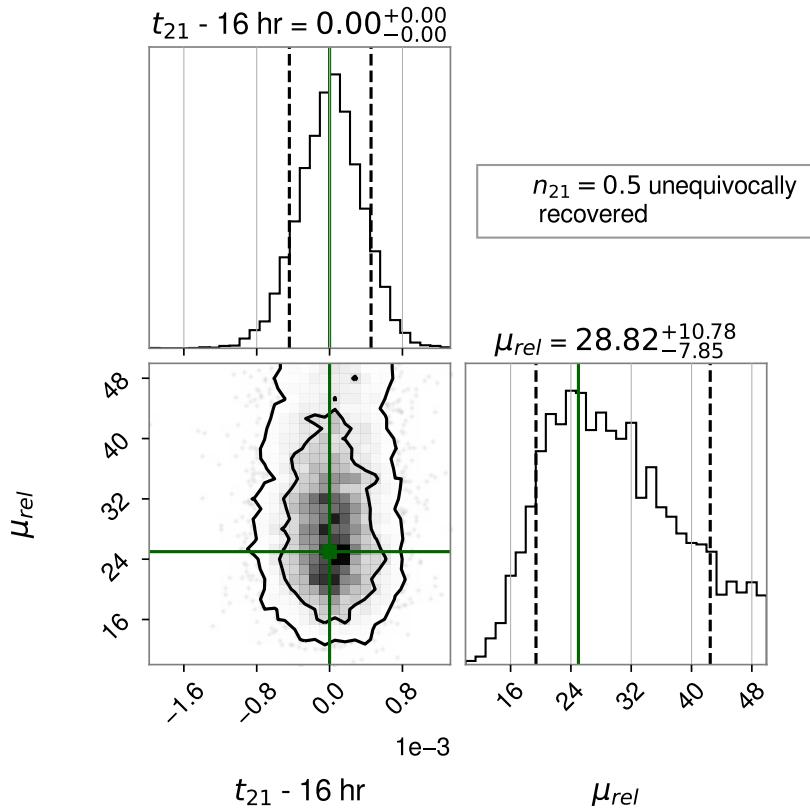


Figure 8.3: Posterior distribution of the magnification, and time delay (re-centered at zero) between two strongly lensed gravitational-wave images, where one of the events is sub-threshold (with an SNR of ~ 5.5). The simulated relative magnification $\mu_{\text{rel}} = 25$, relative Morse factor $n_{21} = 0.5$, and relative time delay $t_{21} = 57600\text{s} = 16\text{ hr}$. The parameters are well recovered, and the Morse factor difference is also uniquely recovered in this scenario.

Eq. (8.13). The most notable impact is on the sky localization, where the 90% confidence sky area improves by about a factor of two in our example case (Fig. 8.2). This is particularly important for the strong lensing science case as an improved sky localization might help narrow down the number of possible host galaxies when combining the GW information with EM data [192].

As an additional example, we analyze a sub-threshold trigger (a signal hidden in the noise background). In a targeted sub-threshold search, one uses a reduced template bank to cover the source parameters posteriors recovered from the primary super-threshold event [254, 255] – such searches may uncover many additional candidates, which would need to be analyzed. We assume that a super-threshold event has already been observed (with identical parameters to the event described in the earlier example). The lensing parameters for the sub-threshold counterpart are $\mu_{\text{rel}} = 25$,³ $n_2 = 1$, and $t_{21} = 16\text{ hr}$, leading to a network SNR of ~ 5.5 , which is below the value typically required for

³For this search, the relative magnification prior is extended to cover the $[0.01, 50]$ interval.

detection.

The resulting posteriors are shown in Fig. 8.3. We recover injected values, but (as expected) the relative magnification measurement is less accurate than it is for typical super-threshold events. The coherence ratio is $\log \mathcal{C}_U^L = 9.3$, consistent with lensing.

8.5 Adaptation to Multiple-Image Analyses

The previous sections have shown how one can use the conditioned likelihood to analyze a pair of events rapidly. However, lensing can lead to more than two images [164]. Our approach can easily be generalized for any number of images while remaining computationally tractable.

If one has N GW lensed images, their detector-frame parameters are related under the lensed hypothesis, while they are unrelated under the unlensed hypothesis. Therefore, the coherence ratio takes the form

$$\mathcal{C}_U^L = \frac{p(d_1, d_2, \dots, d_N | \mathcal{H}_L)}{p(d_1, d_2, \dots, d_N | \mathcal{H}_U)}, \quad (8.17)$$

where the numerator can be written as

$$p(d_1, d_2, \dots, d_N | \mathcal{H}_L) = p(d_1 | \mathcal{H}_L) \prod_{i=2}^N p(d_i | d_1, \dots, d_{i-1}, \mathcal{H}_L), \quad (8.18)$$

and the denominator as

$$p(d_1, d_2, \dots, d_N | \mathcal{H}_U) = \prod_{i=1}^N p(d_i | \mathcal{H}_U). \quad (8.19)$$

The conditioned evidence for the i^{th} image in Eq. (8.18) is computed similarly to the two-image case (8.10):

$$p(d_i | d_1, \dots, d_{i-1}, \mathcal{H}_L) = \int d\phi_i L_i(\phi_i | \mathcal{H}_L) p(\phi_i | \mathcal{H}_L), \quad (8.20)$$

where ϕ_i are the relative lensing parameters for the i^{th} image, i.e. the parameters linking the current image with the first image considered. In Eq. (8.20), the marginalized likelihood is obtained by averaging the likelihood over the reweighed samples coming from the $(i-1)^{th}$ image analysis, hence

$$L_i(\phi_i | \mathcal{H}_L) = \langle p(d_i | \Theta, \phi_i, \mathcal{H}_L) \rangle_{p(\Theta | d_1, \dots, d_{i-1}, \mathcal{H}_L)}, \quad (8.21)$$

where we implicitly include the Morse factors for images 1 to $i-1$ in Θ .

As before, the final posteriors, accounting for the i first images, can be obtained through a reweighing process as

$$p(\Theta, \phi_i | d_1, \dots, d_i) \propto \frac{p(d_i | \Theta, \phi_i)}{p(d_1, \dots, d_i | \phi_i)} p(\Theta | d_1, \dots, d_{i-1}) p(\phi_i | d_1, \dots, d_i). \quad (8.22)$$

In essence, this means that the analyses of additional images after the first two amount to one extra GOLUM run followed by reweighing for each additional image. So, one starts by analyzing the two first images. Then, the third image is analyzed by performing the GOLUM run using the reweighed samples from the first pair analysis. Once the samples obtained for the three-image run are reweighed, they are used as a prior for the GOLUM run of the fourth image. The process is repeated until all images are analyzed. Since only GOLUM runs are involved, the addition of images is relatively cheap.

8.6 Quadruple Image Analysis: Sky Localization

Let us analyze an example quadruplet of lensed images. We assume that the first and second images have the same parameters as in Sec. 8.4. We inject two more lensed signals, with relative magnifications of 4 and 5, time delays of 16 hours and 21 hours, and Morse factors $n_3 = 0$ and $n_4 = 1$, respectively.

We begin by analyzing the first two images. We then use the reweighed samples obtained from the joint analysis of the first two images to analyze the third image. As a consequence, we retrieve the lensing parameters $p(\phi_3 | d_1, d_2, d_3, \mathcal{H}_L)$ and the conditioned evidence $p(d_3 | d_1, d_2, \mathcal{H}_L)$ for the third image. We then reweigh the posterior samples from the second run with the results from the third, obtaining $p(\Theta, \phi_3 | d_1, d_2, d_3)$. Using those reweighed samples, we analyze the fourth image similarly, obtaining $p(\Theta, \phi_4 | d_1, d_2, d_3, d_4)$.

A particularly noteworthy improvement is in the sky localization, which we show in Fig. 8.4. The initial 90% sky area of $\sim 20 \text{ deg}^2$ of the first image is reduced to a final area of $\sim 2 \text{ deg}^2$ when accounting for the four images. Such an improvement is important for studies involving lensed host galaxy localization, which rely on an accurate sky map estimate [192]. The entire analysis was performed in around 4 CPU hours⁴.

⁴This corresponds to the time needed to infer the lensed parameters and reweigh the posteriors for the three images.

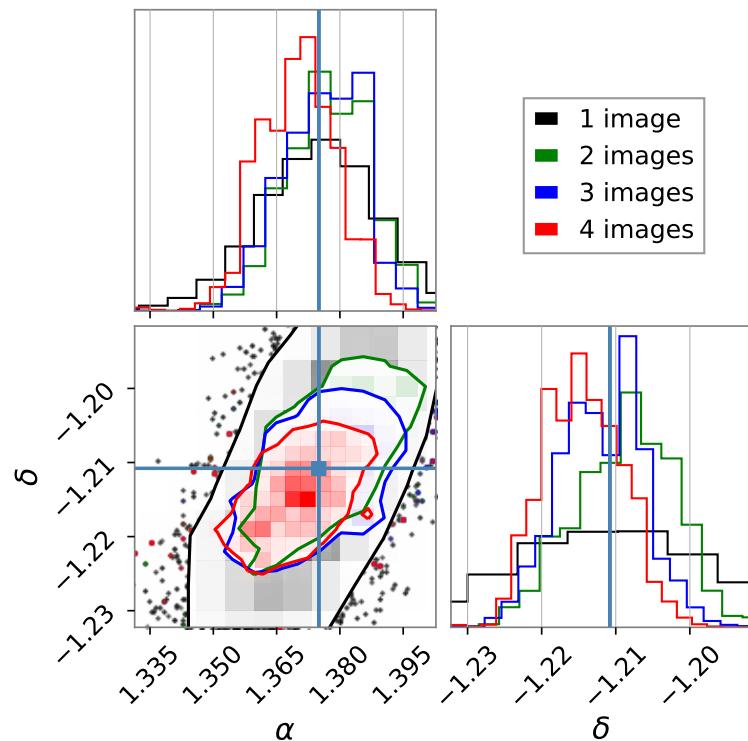


Figure 8.4: The 90% credible region of the sky location of a strongly lensed GW singlet (black), pair (green), triplet (blue), and quadruplet (red). There is a clear improvement in sky localization with every GW image added. The final 90% confidence sky area is $\sim 2 \text{ deg}^2$ in this example. An improved sky localization might be particularly useful for lensed host galaxy localization.

8.7 A Note on Selection Effects

The ratio of evidence (5.32) is called the coherence ratio instead of the Bayes factor because it does not include selection effects. The latter accounts for the selection bias in the observed events due to the detectors' sensitivity [214, 257]. Therefore, one needs to account for the probability of detecting given events in the data, which itself depends on the CBC merger population. The difficulties for this process lie in the uncertainties present in the BBH and lens population [189]. Accounting for selection effects, the evidence under the lensed hypothesis becomes [214]

$$\begin{aligned} p(d_1, d_2 | \text{det}, \mathcal{H}_L) &= \frac{1}{p(\text{det} | \mathcal{H}_L)} \int \int d\Theta d\phi p(d_1, d_2 | \Theta, \phi, \mathcal{H}_L) p(\Theta, \phi | \mathcal{H}_L) \\ &= \frac{p(d_1, d_2 | \mathcal{H}_L)}{p(\text{det} | \mathcal{H}_L)}, \end{aligned} \quad (8.23)$$

where “det” means the data is conditioned on its detection, and $p(\text{det} | \mathcal{H}_L)$ is the probability to detect a signal under the lensed hypothesis

Through a similar derivation, the unlensed hypothesis becomes

$$p(d_1, d_2 | \text{det}, \mathcal{H}_U) = \frac{p(d_1, d_2 | \mathcal{H}_U)}{p(\text{det} | \mathcal{H}_U)^2}. \quad (8.24)$$

Finally, the Bayes factor with selection effects is [214]

$$\mathcal{B}_U^L = \frac{p(\text{det} | \mathcal{H}_U)^2}{p(\text{det} | \mathcal{H}_L)} \frac{p(d_1, d_2 | \mathcal{H}_L)}{p(d_1, d_2 | \mathcal{H}_U)} = \frac{p(\text{det} | \mathcal{H}_U)^2}{p(\text{det} | \mathcal{H}_L)} \mathcal{C}_U^L. \quad (8.25)$$

In practice, GOLUM can be linked to HANABI [214] to use its selection effect capabilities or the selection effects can easily be added to the framework. In essence, it can be done by reweighing the evidence to account for a BBH population and a lens model. The factors in front of the coherence ratio in Eq. (8.25) are computed once since it does not depend on the lensing parameters and only on the populations and detector sensitivities.

8.8 Concluding Remarks Concerning GOLUM

In this chapter, we introduced a fast and accurate methodology to perform strong-lensing analyses. It relies on recasting the joint likelihood as a conditioned likelihood. The first image run amounts to a standard BBH parameter estimation run with an additional Morse factor. The second image run can be expressed as a marginalized likelihood and

further be sped up using a look-up table. Finally, the joint posteriors are obtained using a reweighing process. This setup significantly reduces the computational time compared to joint parameter estimation. Additionally, the method can easily accommodate more images without significantly increasing the computation time. One can also easily add selection effects to the pipeline or link it to existing ones, such as [214].

The combination of speed and precision allowed by our method will likely become crucial in the future when we expect the number of detected individual events, each of which could, in principle, be a lensed image, to rise rapidly.

This framework could also play a complementary role to existing methods such as posterior overlap [181] and joint parameter estimation [214, 215]. It would perform strong lensing estimates and multiple-image analyses in an accelerated fashion – situated between the two existing methodologies in terms of speed and precision. A three-tier analysis may be possible, where we first analyze the strongly lensed events with a posterior-overlap method, after which we analyze a reduced set of events with our methodology, and finally, the best candidate(s) could be passed to the joint parameter estimation tools.

9

ON THE IDENTIFICATION OF INDIVIDUAL GRAVITATIONAL-WAVE IMAGE TYPES OF A LENSED SYSTEM USING HIGHER-ORDER MODES

Identifying strong-lensed event pairs is not an easy task [181, 187, 202]. Therefore, it is of interest to try and find characteristic features related to lensing. This is the case for the Morse factor (see Eq. (5.22) for the link between lensed and unlensed waveforms). This overall phase-shift is not degenerate with other BBH parameters when HOMs are present [169, 170]. Therefore, one can try to identify it. A clear identification would lead to smoking-gun evidence for strong lensing. Past searches have focused on the case of single type II images [115, 116]. In Ref. [115], the authors focus on a single detected image in a 3G detector scenario. In such a case, since the SNR for a BBH is usually very large, many systems have a detectable HOM contribution. On the other hand, in [116], the authors focus on 2G detectors. In addition to inquiring which systems give rise to detectable Morse factors, they study potential biases in the parameter estimation due to the non-inclusion of the Morse factor. For the detection and the bias, only signals with a large inclination and high masses lead to observable effects. Moreover, they require relatively high SNR. However, very asymmetric systems are disfavored by current population models [46] and are also less likely to be observed.

In Ref. [117] and this chapter, we investigate the possibility of detecting type II images when two lensed events are observed. In this case, since we are comparing one signal (a type II image) with a reference signal (a type I image), the identification is made easier, and the Morse factor effects are detectable at lower total HOM contributions.

9.1 Methodology

The goal of this work is to understand the circumstances under which HOMs can enable us to identify the types of images present in a lensed pair. To explore the effect of different HOM contributions, we tune the HOM SNR by varying the mass ratio $q = m_2/m_1$ (with m_1, m_2 the component masses), the inclination ι , and the luminosity distance

D_L of the events. For the first image, D_L is adapted so that the network SNR (see Eq. (4.12)) for the event is always 12. For the mass ratio, the three values considered are 0.1, 0.3, and 0.5, while for the inclination we choose the values to be 20° , 45° , and 70° . When considering lensing, we also need to specify the image types as well as the relative magnification μ_{rel} and time delay t_{21} between the two images. We consider three types of lensed systems: type I–type II, type I–type I, and type II–type II, where a type I image has a Morse factor $n = 0$ and a type II image has $n = 0.5$. One could also have type III images, with $n = 1$. However, those are expected to be rare as they require lenses with very shallow profiles [258–260] and are therefore not considered here. Throughout this work, the time delay between the two images is arbitrarily fixed at 11 hr while the relative magnification is such that the SNR of the second image has a specific value. We consider values of 12 or 25 for the second image as these represent a typical and a loud event based on current LVK observing runs, respectively [40].

For each event, we inject the GW with the IMRPhenomXHM waveform model [261], a non-precessing waveform with HOMS, in a network of interferometers made of the two LIGO detectors and the Virgo detector at design sensitivity [25, 26] assuming Gaussian, stationary noise and we perform the analyses with IMRPhenomXHM and IMRPhenomD [104] as template waveforms. To this effect, we use the GOLUM joint parameter estimation framework laid out in Ref. [191] and Chapter 8 for analyzing multiple lensed images. The main idea behind this framework is to use the posterior from one image as the prior for another image, which together with the use of a lookup table leads to a significant speed-up in the analysis.

The priors used in our analyses and the general setup are the same as in Sec. 6 of [191]. That is, the priors for the lensing parameters are uniform for the relative magnification and time delay and discrete uniform for the (difference in) Morse factor. Furthermore, we choose a uniform prior for the chirp mass, the mass ratio, the coalescence time, the cosine of the inclination angle, the polarization angle, and the coalescence phase. The prior for the sky position is such that we have a uniform distribution for the location on a sphere, and the luminosity distance prior is uniform in comoving volume.

9.2 Results

Here we first look at the possibility of identifying the individual image types for an observed pair of lensed images. We investigate how our ability to do so evolves with the HOM content of the image pair and contrast this with the scenario where only one image is detected. We also look at the impact of analyzing an event pair with HOMs using a waveform that does not include them. Finally, we investigate whether our ability to discern the HOM content (and not only the image type) improves when we analyze two images jointly.

9.2.1 Type I–Type II Systems

First, we consider a system of type I and type II images and investigate our ability to recover the image types; we contrast this with the case of a single type II image. Note that when performing a joint analysis on two images, the *difference* in the Morse phase can always be determined unambiguously for the systems considered (with $n_{21} = 0.5$ for the image configuration at hand). From this information, one can infer straight away that the first image is not a type III image. Next, as a heuristic criterion to determine that an image type is correctly recovered in the two-image case, we choose that the posterior probability $P(n_1 = n_1^{inj} | \text{data}) \geq 0.75$, where n_1^{inj} is the injected value of the Morse factor of the first image. Indeed, when we have no information at all about the image type, we expect a probability of 0.5 for both image types. The value of 0.75 corresponds to half of the probability of the disfavored image type going to the correct one. We find this to be the case, on average, once the HOM SNR (defined as the quadrature sum of the SNRs over individual modes and the two images) satisfies $\rho_{\text{HOM}} \gtrsim 0.5$. On the other hand, for a single image, we cannot immediately discard the type III image scenario. So, there are now three image types to consider. When we have no information about image types, each of these comes with a probability of 0.33. Here we choose $P(n_1 = n_1^{inj} | \text{data}) \geq 0.66$ as the (again heuristic) criterion to determine that the event type is correctly identified in the case of single images; as before, this corresponds to half of the probability of the disfavored image types going to the correct one. This threshold is crossed when $\rho_{\text{HOM}} \gtrsim 1.3$. Consequently, for a lensed event pair, identifying the image types can be done at a weaker HOM contribution than for a single type II image. A comparison of the way $P(n_1 = n_1^{inj} | \text{data})$ evolves with the HOM SNR can be seen in Fig. 9.1. The decision threshold is crossed for a lower ρ_{HOM} when two images are observed.

The unequivocal recovery of the image types for an event pair would constitute smoking-gun evidence for lensing, as no other “standard” effect could reproduce similar results [115, 170].

9.2.2 Type I–Type I and Type II–Type II Systems

Let us now consider other types of systems, namely, type I–type I and type II–type II.

The ability to identify the image types for a given pair depends on the types of images present. As in the previous case, when two type II images are detected, we find that the image types can be identified at a lower total HOM SNR than for the observation of a single type II image.

On the other hand, we cannot identify the image types unequivocally for type I–type I systems, regardless of the HOM SNR. When the HOM SNR is high enough, it is possible

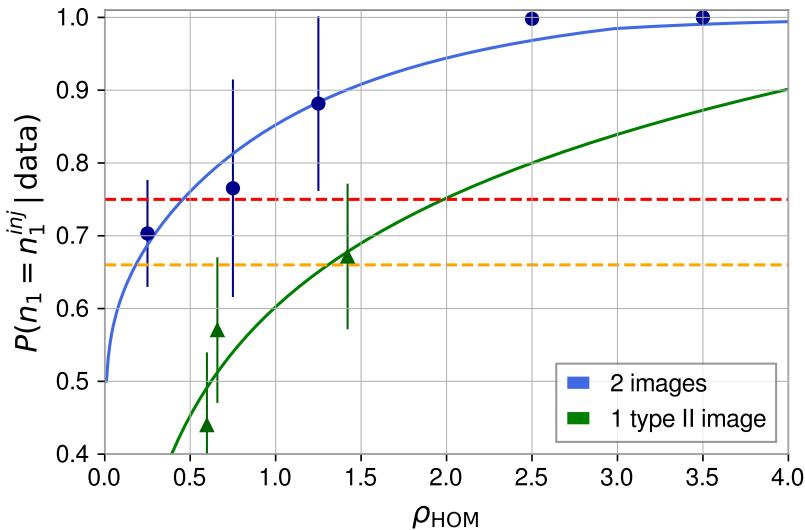


Figure 9.1: Comparison between the posterior probability values for the recovery of the Morse factor, for systems made of a type I and a type II image (in blue), and for a single type II image (in green) as a function of the total SNR in the HOMs. For the lensed system, the first image corresponding to the type I image has a fixed SNR of 12, while the second image has an SNR of 12 or 25. We change the HOM content of the images by using different combinations of mass ratios and inclinations and show medians and 90% intervals for the distribution of probabilities. For the single image systems, the SNR is fixed at 25, and we change the HOM content by varying the mass ratio and the inclination. The image type identification is made at lower total HOM content when two images are observed than when only one type II image is observed. The recovery for the type II image is the same as for the type I image, as the difference between the Morse factor is always unequivocally recovered.

to exclude the presence of a type II image, and we can say that the two images are of the same type. However, it will be difficult to distinguish type I-type I from type III-type III systems based on GW data alone. That said, type III imaging is expected to be rare when considering a galaxy lens [258–260], so in that sense, the interpretation of two type-I images will be preferred. On the other hand, the situation regarding type III is less clear when galaxy cluster lenses are considered.

These observations show that we will likely require at least one type II image to determine the image types based on GW data alone.

9.2.3 Using Templates Without HOMs to Analyze Systems with HOMs

Without HOMs, the coalescence phase and the Morse phase are degenerate; hence image type identification is impossible when using template waveforms without HOMs [170]. In addition, the non-inclusion of the HOM in the analysis of events containing significant

HOMs can lead to biases in some parameters such as the polarization angle, the phase, and the distance [262, 263]. Since this bias will change depending on the antenna response of the detector, the two images making up the lensed system are biased differently. And indeed, for a type I–type II system, when the HOM content is strong (e.g. $\rho_{\text{HOM}} = 3.5$), our framework is not able to detect lensing any longer when the analysis is done with IMRPhenomD.

9.2.4 Improved Probing of HOMs with Lensing

Finally, we compare parameter estimation results for a type I–type II image system with those for a single unlensed image, both having the same total SNR (with a value of 16.97 for $\sqrt{\sum_{i=1,2} \text{SNR}_i^2}$, where i runs over the images) and $\rho_{\text{HOM}}/\rho_{\text{tot}} = 0.14$. We use the same BBH parameters for the different types of systems, except for the polarization angle and the (apparent) luminosity distance. In that sense the total HOM content is the same in both scenarios, enabling us to probe whether observing a lensed pair of events leads to better inference on the HOMs.

As an important example, HOMs allow us to better constrain the orbital inclination, as seen in Fig. 9.2. Hence, detecting two lensed images with the presence of HOMs would allow us to study the HOM content with greater precision. This is likely to have implications, for example, for the use of GW lensing in cosmology [192], or testing GR by probing the polarization content of gravitational waves [200].

9.3 Summary and Conclusions

In this chapter, we have focused on the impact lensing and HOMs can have on each other when observing a lensed image pair. We have shown that our ability to identify the strong lensing image types greatly improves when jointly analyzing two images as opposed to one. If we were to identify the presence of type-II images, it would count as smoking-gun evidence that the event is indeed lensed. In addition, we have confirmed that the presence of a type II image is required to unequivocally identify the observed image types based on GW data alone. We have also shown that when the HOMs play an important role, their non-inclusion in the lensing analysis can lead to the non-detection of a lensed pair. Finally, we have shown that strongly lensed gravitational-wave events allow us to study the HOM content more accurately than similar non-lensed gravitational waves, likely improving several scientific applications for strongly-lensed GWs.

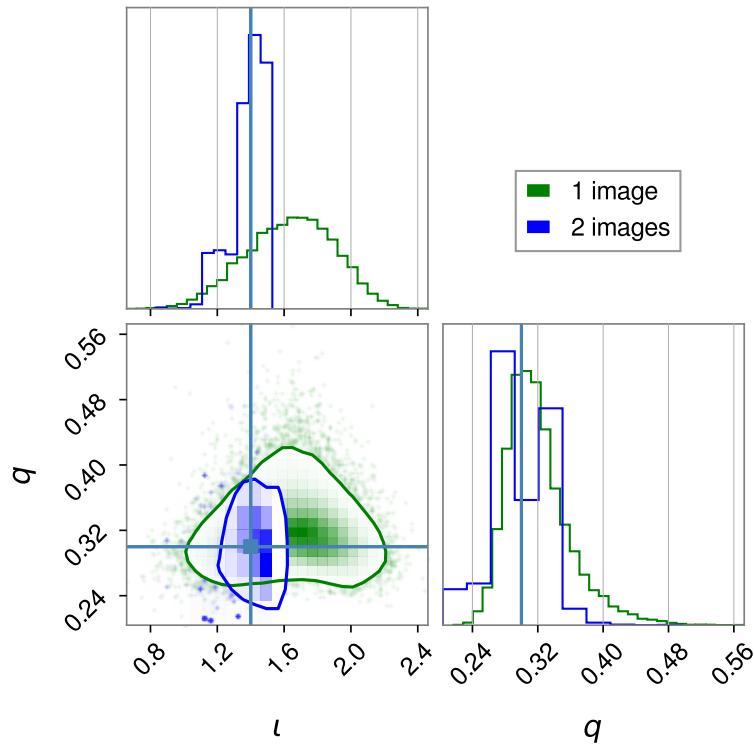


Figure 9.2: The posterior distribution for the inclination (ι) and the mass ratio (q) for an unlensed event (in green) and a lensed image pair (in blue). The events have the same (total) network SNR (of 16.97), and the same $\rho_{\text{HOM}}/\rho_{\text{tot}}$ (of 0.14). Even if the posterior on the mass ratio is not significantly better, the one on the inclination is $\sim 2\times$ narrower. Hence, the support of the posterior in the q - ι plane has a smaller surface for the lensed scenario, showing that the HOMs are better constrained when we observe a lensed image pair than in the case of a single image with the same total HOM content.

10

ORDERING THE CONFUSION: A STUDY OF THE IMPACT OF LENS MODELS ON GRAVITATIONAL-WAVE LENSING DETECTION CAPABILITIES

Searches for strong lensing are essentially looking for events with the same frequency evolution. However, there are uncertainties about the measured GWs' characteristics. Therefore, when too many events are present in the data, the FAP increases rapidly [187, 201, 264].

However, one can have some expectations about the lens. For example, one can think the lens is most likely a galaxy. In this case, it is possible to compute expected distributions for the lensing parameters [181, 187, 203]. From this, one computes *lensing statistics*, representing how likely it is to observe the (apparent) lensing parameters under the lensed and unlensed hypotheses; see Eq. (5.36).

In some previous works [181, 187], it has been shown how using this ratio in parallel with posterior overlap can reduce the FAP. In Ref. [202] and this chapter, we go a step further by working this statistic into the coherence ratio in a post-processing step, enabling one to test several models without needing to redo the nested sampling run several times. We then proceed by showing how using a more precise method such as GOLUM decreases the FAP, and how the coherence ratio including information about the lens makes for a better discriminator when performing lensing searches. We also show the risk of taking the wrong model, where one can worsen the situation and increase the FAP.

10.1 Including the Lens Statistics in the Coherence Ratio

The coherence ratio as presented in Eq. (5.32) accounts for similar frequency evolution of the signals but, generally, does not assume anything particular on the lensing parameters. So, searches use uninformative priors on the lensing parameters to not bias their results towards a pre-defined model. However, one may want to constrain the lensing parameters given some models. This can be done by using adapted priors. However, doing this means

one has to redo the parameter estimation runs for each model considered. Therefore, it is more convenient to constrain the lensing parameters after the analysis is done.

The inclusion of a lens model in the coherence ratio can be done as follows. If $\mathcal{Z}_{\mathcal{H}_I}^M$ is the evidence for a given model M under the hypothesis \mathcal{H}_I , and $\mathcal{Z}_{\mathcal{H}_I}^A$ is the evidence obtained from the model-agnostic analysis for the same hypothesis, then

$$\mathcal{Z}_{\mathcal{H}_I}^M = \left\langle \frac{p(\boldsymbol{\vartheta}_I | M, \mathcal{H}_I)}{p(\boldsymbol{\vartheta}_I | A, \mathcal{H}_I)} \right\rangle_{p(\boldsymbol{\vartheta}_I | D, A, \mathcal{H}_I)} \mathcal{Z}_{\mathcal{H}_I}^A. \quad (10.1)$$

In this expression, $\boldsymbol{\vartheta}_I$ represents the set of all parameters needed to describe the data under hypothesis I , $p(\boldsymbol{\vartheta}_I | M, \mathcal{H}_I)$ and $p(\boldsymbol{\vartheta}_I | A, \mathcal{H}_I)$ are the probabilities of observing the parameters in the model and the analysis for a given hypothesis. $p(\boldsymbol{\vartheta}_I | D, A, \mathcal{H}_I)$ is the posterior distribution obtained from the model-agnostic analysis for data D .

Indeed, for a given hypothesis (not written explicitly here to ease the notation), the evidence for a model M is

$$\mathcal{Z}^M = \int d\boldsymbol{\vartheta} p(D|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta}|M), \quad (10.2)$$

where D is the data, which can be composed of several data streams. However, using Bayes' theorem, the posterior for another situation A is

$$p(\boldsymbol{\vartheta}, | D, A) = \frac{p(D|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta}|A)}{p(D|A)} = \frac{p(D|\boldsymbol{\vartheta})p(\boldsymbol{\vartheta}|A)}{\mathcal{Z}^A}. \quad (10.3)$$

So,

$$p(D|\boldsymbol{\vartheta}) = \frac{\mathcal{Z}^A p(\boldsymbol{\vartheta}|D, A)}{p(\boldsymbol{\vartheta}|A)}. \quad (10.4)$$

Combining Eqs. (10.2) and (10.4), one gets

$$\begin{aligned} \mathcal{Z}^M &= \int d\boldsymbol{\vartheta} \mathcal{Z}^A \frac{p(\boldsymbol{\vartheta}|D, A)}{p(\boldsymbol{\vartheta}|A)} p(\boldsymbol{\vartheta}|M) \\ &= \mathcal{Z}^A \int d\boldsymbol{\vartheta} \frac{p(\boldsymbol{\vartheta}|M)}{p(\boldsymbol{\vartheta}|A)} p(\boldsymbol{\vartheta}|D, M) \\ &= \left\langle \frac{p(\boldsymbol{\vartheta}|M)}{p(\boldsymbol{\vartheta}|A)} \right\rangle_{p(\boldsymbol{\vartheta}|D, A)} \mathcal{Z}^A. \end{aligned} \quad (10.5)$$

The last line corresponds to Eq. (10.1), as announced.

In practice, one does not solve the integral over the ratio of probabilities but uses the samples obtained from the analyses to compute the weights for each set of samples and

then takes the average (hence performing a Monte Carlo integration). So,

$$\mathcal{Z}_{\mathcal{H}_I}^M = \frac{1}{N} \left(\sum_{i=0}^{i=N} W_A^M(\vartheta_I^i, \mathcal{H}_I) \right) \mathcal{Z}_{\mathcal{H}_I}^A, \quad (10.6)$$

where N is the total number of samples, and

$$W_A^M(\vartheta_I^i, \mathcal{H}_I) = \frac{p(\vartheta_I^i | M, \mathcal{H}_I)}{p(\vartheta_I^i | A, \mathcal{H}_I)} \quad (10.7)$$

is the ratio of probabilities between the model and the analysis for a set of parameters i . Here, the $\{\vartheta_i\}_{i=1,\dots,N}$ samples are drawn from $p(\vartheta_I | D, A, \mathcal{H}_I)$.

The model-dependent coherence ratio is then obtained by taking the ratio of the evidence for the lensed and the unlensed hypotheses for the model:

$$\mathcal{C}_U^L \Big|_{Model} = \frac{\mathcal{Z}_{\mathcal{H}_L}^M}{\mathcal{Z}_{\mathcal{H}_U}^M}. \quad (10.8)$$

In the end, since the reweighing process is much faster than the parameter estimation analysis, this approach enables one to adapt the results for different models without significantly increasing the computational burden. In addition, if the initial coherence ratio is low, one already knows the event is not lensed since the parameters should match regardless of the lens model. So, the model-dependent part of the analysis is not needed. A good strategy would be to first carry out the parameter estimation for all the events and then apply the reweighing to account for the effect of the various lens models for the events with a high coherence ratio.

10.2 Injections and Setup of the Study

10.2.1 Binary Black Hole Population

Here, we study the impact of the lens model included in the coherence ratio computation on our ability to differentiate between lensed and unlensed events. Since the fraction of strongly-lensed events is relatively low, $\mathcal{O}(10^{-3})$ [e.g., 187, 265], we focus on making a large unlensed background with lensed events on top. Therefore, we generate 100 unlensed BBH mergers. Their masses are sampled from the PowerLaw + Peak distribution [46]^{1 2}.

¹For all the models in this study, we took the median parameter values found in [266].

²The PowerLaw + Peak model, as well as the other models shown in [46], present a secondary peak around $30 M_\odot$. This can lead to more events with closer masses and an increased FAP for lensing.

The spins and redshifts are sampled from the ones observed by the LVK collaboration [46]. The sky location is sampled from a uniform distribution over the sky, the inclination is uniform in cosine, and the phase and polarization are uniform in their domain. We take the time of arrival for the unlensed events to be uniform in a year. For each event, we draw randomly from one of the following cases - the event is observed by i) the two LIGO detectors ii) one of the LIGO detectors and the Virgo detector, or iii) by the three detectors jointly. It is important to vary the number of detectors since fewer detectors lead to larger uncertainty on some critical parameters, such as the sky location. In turn, it leads to more overlap between the posteriors and higher coherence ratios for unlensed events. For each parameter set drawn from the distribution, we take the PSD to be that of one of the events in GWTC-2.1 [267, 268] or GWTC-3 [40, 269] and generate colored Gaussian noise from it. We then inject the GW strain into the noise. This leads to a realistic scenario for detections where the number of detectors and the noise realization are different from one event to the other³. The change in the observation conditions between events is important as the differences in noise and number of detectors change the accuracy we have from one event to the other, impacting the observed detection statistic.

From these 100 unlensed events, we make 4950 unlensed pairs. We also add 50 lensed event pairs. For the first image, the masses, spins, and apparent luminosity distances are drawn from the same distributions. We then generate the second image by selecting the relative magnification and the time delay from the $\mathcal{M}_{\mu,t}$ parameter catalog [203]. From here on, we take these distributions as the true lensing parameter distribution.

We analyze the different events under the unlensed hypothesis using BILBY [149], and the events under the lensed hypothesis using the GOLUM framework [191] which provides fast and accurate joint parameter estimation for strong lensing. The two parameter estimation pipelines are used with the DYNESTY sampler [151].

10.2.2 Population Analyses

Using our background, we perform different analyses to understand the process of identifying the lensed events in an unlensed background better. For each event pair, we perform a posterior overlap analysis [181]⁴ as well as a joint parameter estimation run [191]. Since the posterior overlap method considers only a subset of the posteriors, while GOLUM requires the strains to be consistent, this is a way to check the effect of using a more accurate parameter estimation tool.

³Here, the realistic observation conditions apply to the changes in the number of detectors and noise from one event to the other. The proportion of lensed and unlensed events is not realistic, but the high number of lensed events is needed to define the threshold for the FAP.

⁴The consistency between posteriors is computed for the component masses, the sky location, the spin amplitudes, and tilt angles, and the binary's inclination, similarly to the approach followed in [172, 188].

In a second step, using the coherence ratios obtained from the joint parameter estimation runs, we reweight them for several models using the procedure explained in Sec. 10.1. We use data from three different catalogs for the lensed models: the \mathcal{H}_t time-delay distribution [181], the $\mathcal{M}_{\mu,t}$ time-delay and relative magnification distribution [203], and the $\mathcal{W}_{\mu,t}$ time-delay and relative magnification distribution [187]⁵. For $\mathcal{M}_{\mu,t}$ and $\mathcal{W}_{\mu,t}$, we do the analysis once with the two lensing parameters included and once with only the time delay. This enables us to probe the impact due to the addition of the relative magnification. A comparison of the time delays and relative magnification distributions for the three models is given in Fig. 10.1.

We also introduce four artificial models which represent various observation scenarios. We denote these models A, B, C, and D. Model A is constructed as a fake galaxy-cluster lens catalog, where we focus on larger relative magnifications and time delays. The two lensing parameters follow a scaled beta distribution:

$$\mathcal{B}(x, p, t, m, M) = m + (M - m)\beta(x, c_1, c_2), \quad (10.9)$$

where $\beta(x, c_1, c_2)$ is the usual beta distribution with domain 0 to 1, and

$$\begin{aligned} c_1 &= 1 + t \frac{p - m}{M - m} \\ c_2 &= 1 + t \frac{M - p}{M - m}. \end{aligned}$$

In this expression, M and m are the maximum and minimum of the distribution, p is its peak, and t is a shape factor. For the relative magnification in model A, the distribution peaks at 10 and has minimum and maximum values of 2 and 30, respectively. The shape factor is 5. For the time delay distribution, the peak is at 3 months, t is 5, and the minimum and maximum values are 3 days and 1 year, respectively. Model B uses the same relative magnification distribution as the $\mathcal{M}_{\mu,t}$ catalog but has a different time delay distribution. We take it as a Gaussian peaking at 4 months with a standard deviation of 1.5 months. This example illustrates the impact of a mismodeling of one of the two parameters. The last two models (C and D) resemble the $\mathcal{M}_{\mu,t}$ model as they focus on the same region of parameters space. However, model C has loose bounds, with $\mu_{ji} \in [0.02, 32]$ and $t_{ji} \in [30 \text{ s}, 400 \text{ day}]$, while model D has tighter bounds, with $\mu_{ji} \in [0.5, 3]$ and $t_{ji} \in [2 \text{ hr}, 60 \text{ day}]$. These two models represent what would happen if one uses tight or loose bounds on the lensing parameters to be more conservative or detect more events respectively.

For each of these models, the probability density in the (μ_{ji}, t_{ji}) -plane is obtained by sampling from the distributions for the individual parameters and performing a KDE reconstruction. The consequence of this is mainly to smoothen the edges of the

⁵We used the code base from [187], but adapted the detector networks and their sensitivity to match our situation.

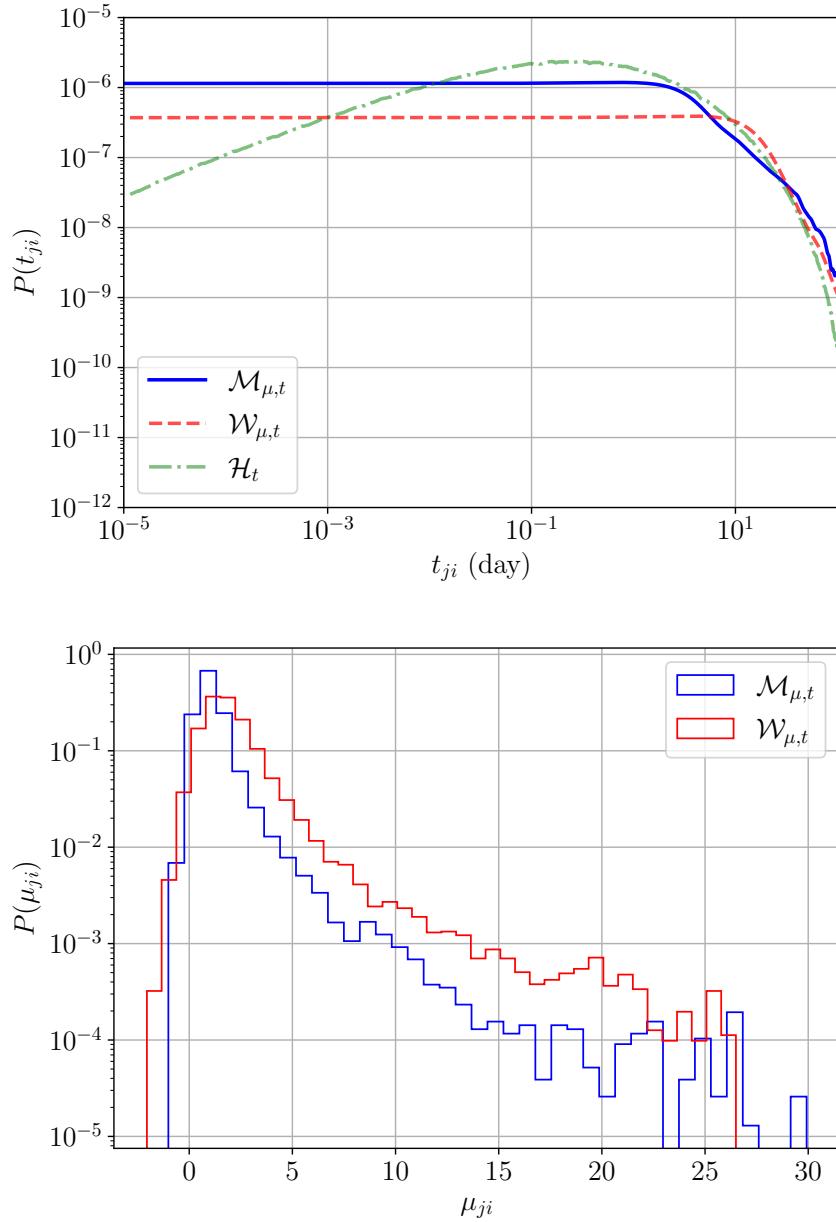


Figure 10.1: *Top:* Time delay probabilities for the different models. One sees that the SIS model (\mathcal{H}_t) peaks at lower values before dropping fast when going to higher values. The SIE-based models have a wider probability distribution. *Bottom:* Relative magnification distributions for the two SIE-based models. The two distributions are consistent, and one sees that adding shear leads to a broader distribution.

Model	Description of the model
$\mathcal{M}_{\mu,t}$	SIE-based model for relative magnifications and time delays described in [203].
\mathcal{M}_t	Same as $\mathcal{M}_{\mu,t}$ but where we only consider the time delay distribution.
$\mathcal{W}_{\mu,t}$	SIE + shear-based model for the relative magnifications and time delays described in [187].
\mathcal{W}_t	Same as $\mathcal{W}_{\mu,t}$ but where we only consider the time delay distribution.
\mathcal{H}_t	SIS-based model for the time delay described in [181].
Model A	Toy model representing galaxy-cluster lenses with scaled beta distributions for relative magnification (peak at 10 with minimum of 2 and maximum of 30) and for time-delay (peak at 3 months, minimum of 3 days and maximum of one year).
Model B	Toy model with same μ_{ji} distribution as $\mathcal{M}_{\mu,t}$ but with a shifted time delay ($\mathcal{G}(4 \text{ month}, 1.5 \text{ month})$).
Model C	Toy model based on $\mathcal{M}_{\mu,t}$ but using broader bounds on the lensing parameters, with $\mu_{ji} \in [0.02, 32]$ and $t_{ji} \in [30 \text{ s}, 400 \text{ day}]$.
Model D	Toy model based on $\mathcal{M}_{\mu,t}$ but using tighter bounds on the lensing parameters, with $\mu_{ji} \in [0.5, 3]$ and $t_{ji} \in [2 \text{ hr}, 60 \text{ day}]$.

Table 10.1: Summary of the different lens models used in this study.

distributions. A summary of the various lens models used in this work is given in Table 10.1.

For the μ_{ji} and t_{ji} distributions of the unlensed events, we use the distributions given in [203] for all of the models except for \mathcal{H}_t . These distributions correspond to a census of magnification (i.e. distance) ratios and time delays for the unlensed pairs made from the BBH population and depend on the specific assumptions of the BBH population. For the \mathcal{H}_t scenario, we use the same approach as in [181], where the times of arrival of individual unlensed events follow a Poisson process. For the unlensed cases in [203] and in [181], the probability density is higher for longer time delays when compared to the lensed scenario (see Fig. 2 in [203] and Fig. 2 in [181] for a representation).

10.2.3 Determining Lensed Candidates

To determine whether events are lensed, we need to use some threshold on the detection statistics. One way of doing this is to use a fixed threshold. For example, one could claim an event to be a lensed candidate if $\ln(\mathcal{C}) > 2$, where \mathcal{C} is any detection statistic. This is similar to the approach considered in [270]. However, this is a generic approach and does not account for the characteristics of the data we are considering. Unlensed events could also cross this threshold by chance, leading to false alarms. Instead, in this

work, we take an approach similar to the one used in [201] with the FAP⁶ given by

$$\text{FAP} = \frac{\#\text{Unlensed} > X}{\#\text{All Unlensed}}. \quad (10.10)$$

Here the numerator is the number of unlensed pairs having their detection statistic above X , a threshold defined based on the observations in the lensed scenario, and the denominator is the total number of unlensed event pairs.

In this work, X is the 5th percentile of the detection statistic for the lensed foreground. Using this method, we can fold in the effect of the models on both the lensed and the unlensed populations. For example, if the model favors the unlensed events and disfavors the lensed ones, its impact on the statistics is such that their values increase for the unlensed events. On the other hand, they decrease for the lensed events, leading to a smaller value of the fifth percentile. Therefore, $\#\text{Unlensed} > X$ increases, and the FAP increases.

In this work, other information is also used to characterize the performance of the detection statistic for a given model. The receiver operator characteristics (ROC) curve represents the ability of the model to differentiate between lensed and unlensed pairs. It represents the efficiency (the fraction of lensed events passing the detection threshold) versus the false-positive probability (FPP; the fraction of unlensed events passing the detection threshold). So, one wants the highest possible efficiency for the lowest possible FPP.

Another way to represent the performance is to use the complementary cumulative distribution function (CCDF) of the unlensed background with the cumulative density function (CDF) of the lensed foreground. Ideally, one wants the CCDF to drop as fast as possible, while the CDF for the lensed foreground should be significant at values as high as possible. The overlap between those two curves will represent the confusion region, where the value of the detection statistic is such that it can correspond to lensed and unlensed events. The smaller this region, the easier it is to distinguish between the two hypotheses.

10.3 Results

In this section, we detail the results obtained in different conditions. First, we compare the FAP between approximate and more precise methods. Then, we look at the effects of model inclusions, going from looking at the impact of the correct model to that of the toy models.

⁶In [201], this is called the FAP per pair, as it corresponds to the probability that a given event pair is classified as lensed while being unlensed.

10.3.1 From Posterior Overlap to Joint Parameter Estimation

First, we verify how moving from posterior overlap to joint parameter estimation modifies the distribution of the corresponding detection statistics. For that, we compute the overlap between the parameters for all the events in our catalog (lensed and unlensed) using the method from [181] as well as the coherence ratio using GOLUM [191].

The comparison between the two is given in Fig. 10.2, where a ROC curve is shown (left panel) as well as a scatter plot (right panel) of the detection statistic for the two methods. These plots show there is a real gain in using a framework like GOLUM, where one ascertains more stringently the correlation between the signals. Based on the ROC curve, the FPP for a given efficiency is reduced when going from one framework to the other. This is also evident from the lower number of unlensed events with $\ln(\mathcal{C}_U^L) > 0$ for GOLUM compared to the posterior overlap. For the lensed events, the two frameworks agree relatively well. Taking the lensing detection threshold as the 5th percentile of the lensed detection statistic distribution, FAP = 0.75% for GOLUM and FAP = 2.9% for posterior overlap, showing that seeking for better correlation between the parameters of the GW signals leads to a lower risk of misidentifying an unlensed event as a lensed one.

Based on this observation, one could advocate using a fast joint parameter estimation tool such as GOLUM to filter out the events before doing more extensive searches. Usually, a GOLUM run would still require the full analysis of the first image, including the Morse factor information. This corresponds to a standard parameter estimation run and is relatively expensive. However, the inclusion of the image types becomes important when there is a strong HOM contribution in the observed event [115–117, 170]. So, a preliminary strategy could be to use the posteriors obtained by the usual LVK pipelines [40], such as BILBY [149], and perform the analysis of the second image using those posteriors, by-passing the more computationally costly first image run. Under the assumption that the HOMs are weak, the distributed coherence ratio [191]

$$\mathcal{C}_U^L = \frac{p(d_1|\mathcal{H}_L)}{p(d_1|\mathcal{H}_U)} \frac{p(d_2|d_1, \mathcal{H}_L)}{p(d_2|\mathcal{H}_U)} \quad (10.11)$$

can be approximated by

$$\mathcal{C}_U^L \simeq \frac{p(d_2|d_1, \mathcal{H}_L)}{p(d_2|\mathcal{H}_U)} \quad (10.12)$$

as the ratio of evidence for the first image is $\mathcal{O}(1)$.

With this method, only the second image run would be needed in GOLUM, and it would take $\mathcal{O}(30 \text{ min})$ at most while enabling a better reduction of the background. Of course, if an event is flagged with a significant HOM contribution, one would necessarily have to redo the joint parameter estimation completely to make sure that nothing was

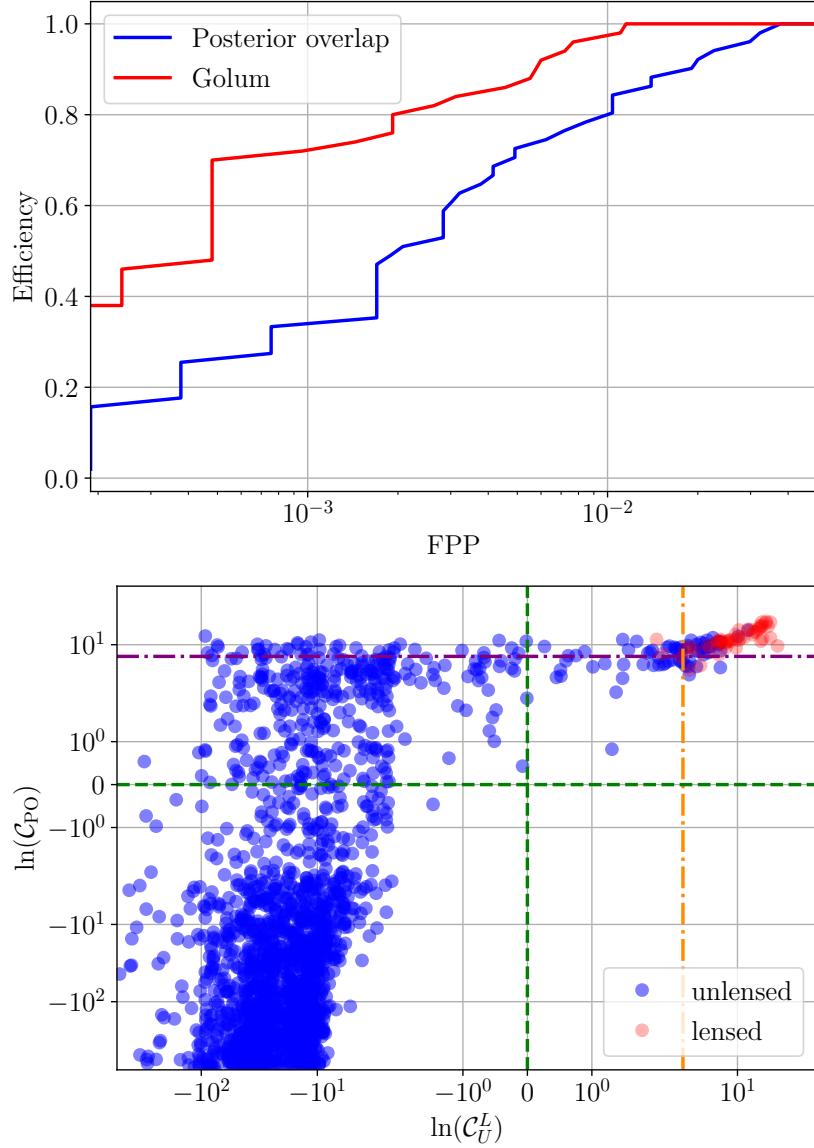


Figure 10.2: *Left:* ROC curves for the GOLUM and posterior overlap methods. Since the curve for GOLUM is more to the left and reaches one faster, it is better suited to determining whether events are lensed. *Right:* Comparison between the $\ln(C_U^L)$ and the $\ln(C_{PO})$ statistics for the lensed and unlensed events in our catalog. The horizontal purple dash-dotted line is the 5th percentile for the lensed events in the posterior overlap analysis, while the vertical orange dash-dotted line is the same quantity for the lensed events analyzed with GOLUM. We see that for the lensed events, the two methods seem correlated. However, there is a clear reduction in the number of high-significance unlensed pairs when using GOLUM .

missed because of potential biases [116, 117]⁷.

Some preliminary investigations performed on our catalog show that most of the events are well recovered without accounting for the Morse factor. More extensive comparisons are left for future work.

10.3.2 Including the Correct Model

Once the catalog has been analyzed and a coherence ratio obtained for all the pairs, one can include the effect of lensing statistic in the final results using Eq.(10.1). This should decrease the confusion region where the background and foreground overlap, and hence the FAP [181, 187, 201]. In this section, we analyze what happens when the *true* model is used. In our case, this means the $\mathcal{M}_{\mu,t}$ model. We denote the detection statistic associated with the $\mathcal{M}_{\mu,t}$ model $\mathcal{C}_{\mathcal{M}_{\mu,t}}$.

A comparison between the background CCDF and the foreground CDF is shown in Fig. 10.3. One sees that the overlap region between the lensed and unlensed distributions is reduced when including the lensing statistics. Indeed, the crossing between the CCDF of the unlensed events and the CDF of the lensed events happens for a higher value and encompasses a smaller area. The unlensed background is decreased for the higher values of \mathcal{C}_U^L , but the tail is not pushed back entirely. This happens because a small number of unlensed events is promoted to a higher $\mathcal{C}_{\mathcal{M}_{\mu,t}}$ when the $\mathcal{M}_{\mu,t}$ information is added. Indeed, amongst the events starting with $\ln(\mathcal{C}_U^L)$ close to zero, some have apparent relative magnifications (i.e. their distance ratios) and time delays more compatible with the lensing hypothesis than the unlensed hypothesis purely by chance. As a consequence, those are not pushed to a lower value but a slightly higher one, mimicking the lensed scenario. However, such events are rare and there is an effective decrease in the number of unlensed events with a high significance. For instance, we go from an FAP = 0.75% for the \mathcal{C}_U^L to an FAP = 0.07% for the $\mathcal{C}_{\mathcal{M}_{\mu,t}}$ statistic. In the end, this confirms that the inclusion of the lensing statistics helps decrease the FAP and makes for more confident detections. This is consistent with previous studies [181, 187, 201, 203].

We note that our values seem a bit more pessimistic than those presented in Ref. [201] because of the following two main reasons. The first one is the number of events analyzed in this work. Indeed, since we need to perform parameter estimation on all of the events and pairs, we do not consider as many events as analyzed in [201]. However, the goals of our works are different and yet complementary. In [201], the goal was to show how difficult it is to identify genuinely lensed pairs in a large number of samples and to show how the FAP evolves with the number of samples. Our goal is to study the effect of the addition of a specific lens model in identifying lensed pairs in an unlensed background. Secondly, we consider more realistic and complex observational conditions. We use

⁷We note that if the HOM content is strong enough to significantly bias the GOLUM analysis, it would also bias the posterior overlap analysis, where the samples are usually taken from a standard unlensed parameter estimation runs.

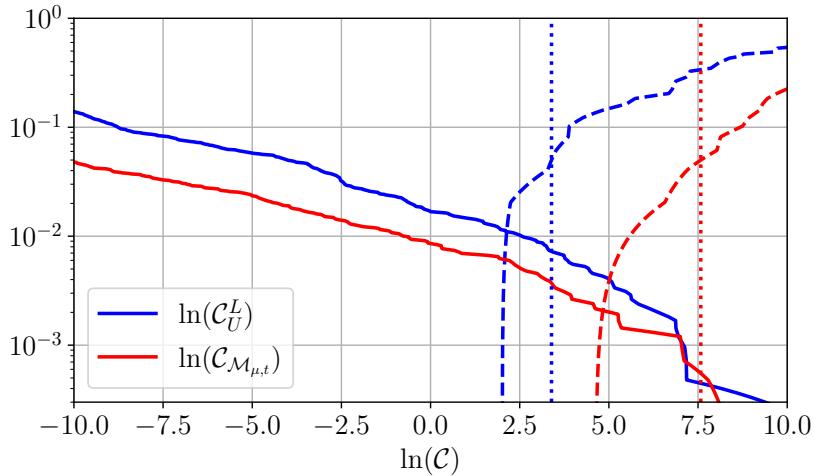


Figure 10.3: Comparison of the CCDF for the unlensed events (solid lines) and the CDF for the lensed events (dashed lines) for the \mathcal{C}_U^L (blue) and the $\mathcal{C}_{M_{\mu,t}}$ (red) statistics. The vertical dotted lines represent the 5th percentile of the lensed distribution, which can be seen as the threshold above which one can consider an event to be lensed. We see that the fraction of unlensed events with a statistic higher than this threshold is reduced significantly when including the lensing statistics.

PSDs coming directly from the third observation run and vary the number of detectors observing different events. This leads to worse constraints on some parameters and more room for matching unlensed events by chance. Nevertheless, both studies suggest that it is difficult to identify lensed pairs in a background of unlensed events, even if the inclusion of a lens model can help in reducing the FAP.

10.3.3 Using Other Models

In the previous section, we have shown that including the expected distributions for the relative magnification and the time delay in the detection statistic helps disentangle the unlensed background and the lensed foreground. However, here, we use the underlying model used to generate the lensed events. When performing real lensing searches, the lens population characteristics are not known accurately (the lens properties for a galaxy-scale or a cluster-scale lens are very different [180, 182–184, 271]). In addition, the models for a given type of lens can also be different. For example, several types of density profiles exist for a galaxy lens, such as SIS [207], SIE [209], and SPEMD [272]. Although some are favored by electromagnetic observations [209], there is no guarantee that the assumed model is the best representation of the true lens population in the Universe, and even the best-fitting models are subject to simplifications and inherent degeneracies. For example, we know that our prediction of the relative magnification is less robust

than the one for the time delay. The former can be impacted by smaller objects present in the macro-lens [e.g., 173, 273], which could lead to smeared distributions or secondary peaks. Therefore, we look at what happens when we use a different lensing statistic catalog and when we use only the time delay coming from the lensing statistics.

10.3.3.1 Effect of Shear

Here, we focus on the difference in detection statistics when including shear in the model while the underlying model has no shear, comparing the results from the $\mathcal{M}_{\mu,t}$ and $\mathcal{W}_{\mu,t}$ models. We note the detection statistic based on the $\mathcal{W}_{\mu,t}$ model $\mathcal{C}_{\mathcal{W}_{\mu,t}}$. As shown in Sec. 10.1 and Fig. 10.1, the two have relatively close distributions, and the main effect of shear is to widen the relative magnification distribution. We also note that the $\mathcal{W}_{\mu,t}$ statistic has a slightly higher probability for large time delays.

A comparison of the detection statistics for the unlensed and lensed events for the two models is shown in Fig. 10.4. One sees the two statistics agree quite well, with a few exceptions. For the most part, the unlensed background is unchanged between the two in the sense that most of the unlensed events for one model are also categorized as unlensed for the other model. For $\mathcal{C}_{\mathcal{W}_{\mu,t}}$, there are a few events with a low $\mathcal{C}_{\mathcal{M}_{\mu,t}}$ pushed to higher statistical values. This happens when the time delays are close to some hundred days and the relative magnification is large. Indeed, in that case, one is in the highest probability values of the model and there is a boost due to the lensing statistics. It is also seen for the $\mathcal{C}_{\mathcal{M}_{\mu,t}}$ statistic where a few events are promoted. This happens for events with short time delays and magnifications close to 1. Still, only a few of the events are significantly changed.

For the lensed events, we see that a few have $\mathcal{C}_{\mathcal{W}_{\mu,t}} < \mathcal{C}_{\mathcal{M}_{\mu,t}}$. This is because the peak probability density is reached for different values of the time delay and the relative magnification. Still, no lensed event is discarded. This decrease in significance for some events leads to an increase in the FAP, as the 5th percentile has a lower value. So, more unlensed events have their value above the threshold. For the $\mathcal{C}_{\mathcal{W}_{\mu,t}}$ statistic, FAP = 0.095%. As a consequence, doing the same analysis with the same density profile as the underlying distribution but with slight variations in the model is still better than using no model at all. Indeed, the FAP is reduced significantly for the $\mathcal{W}_{\mu,t}$ model compared to the statistics for the coherence ratio. This means that small systematics in the lens model should not hinder the benefits coming from the inclusion of a lens model.

10.3.4 Using Only the Time Delay

As mentioned previously, the time delays are less sensitive to systematics for a given model compared to the relative magnification. Therefore, it can be appealing to use

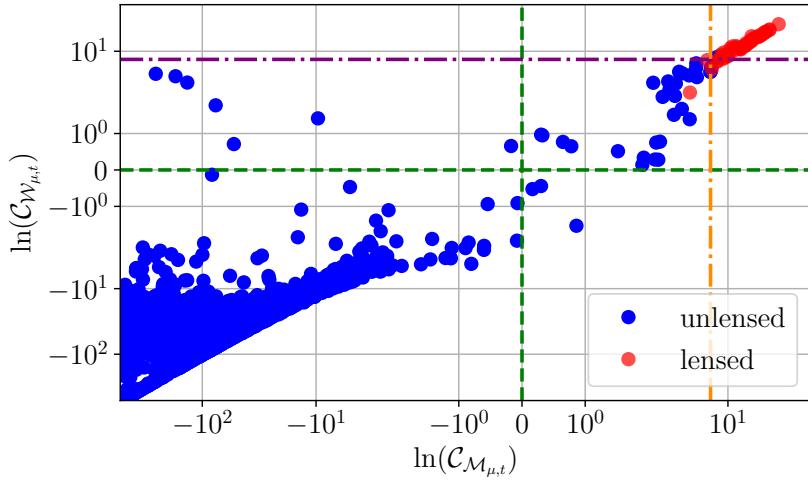


Figure 10.4: Comparison of the foreground and background values for the $\mathcal{C}_{M_{\mu,t}}$ and $\mathcal{C}_{W_{\mu,t}}$ detection statistics. The orange dash-dotted line is the value of the 5th percentile for the $\mathcal{C}_{M_{\mu,t}}$ statistic for the lensed pairs, while the purple one is the same quantity for the $\mathcal{C}_{W_{\mu,t}}$ statistic. For most events, the two models agree reasonably well. Some events have a higher significance for one model or the other. This happens when the apparent lensing parameters are in a high probability region of the model. For the lensed events, the two models agree well, even if there is a slight decrease in significance for some events when using the $\mathcal{W}_{\mu,t}$ distributions. Because of that, the 5th percentile for the lensed event will decrease a bit and the FAP is slightly increased for the $\mathcal{W}_{\mu,t}$ model. However, the difference is small enough to not impact significantly the benefit of including a lens model in the detection statistic.

only the time delay to reweigh the coherence ratio.

To investigate this, we analyze the lensed and unlensed event pairs using the time delay distributions obtained for the $\mathcal{M}_{\mu,t}$ and $\mathcal{W}_{\mu,t}$ models and note these detection statistics \mathcal{C}_{M_t} and \mathcal{C}_{W_t} , respectively. In addition, to study the impact of an error on the lens density profile, we include the time delays obtained from the \mathcal{H}_t model. In this model, the lens profile is an SIS instead of an SIE, leading to a different shape of the time delay distribution (see Fig. 10.1).

A comparison of the performances for the three models, the \mathcal{C}_U^L and the $\mathcal{M}_{\mu,t}$ model is shown in Fig. 10.5. There is no major difference between using the time delay for the SIE or the SIE + shear models, with a small difference at lower FPP, which comes from the lower probability for lower time delays. Still, we see that in this case, the difference between the models is smaller than for the one with the relative magnification included. The two models are slightly less efficient than the correct model including the relative magnification and the time delay. For the $\mathcal{M}_{\mu,t}$ model, some events have a compatible time delay but not a compatible relative magnification. Therefore, they have a lower $\mathcal{C}_{M_{\mu,t}}$. Those events are not flagged here and thus increase the background FPP. A comparison between the $\mathcal{M}_{\mu,t}$ model with and without magnification and the coherence ratio in terms of CDFs and CCDFs for the background and foreground is

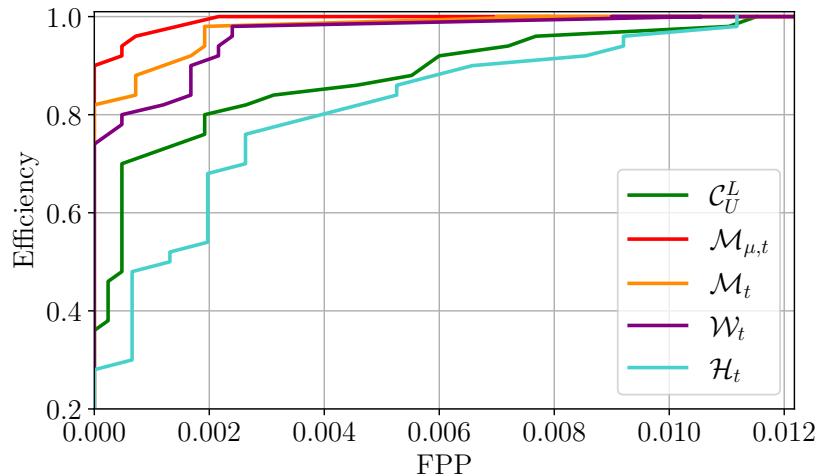


Figure 10.5: ROC curves for the different models with only the time delay included. The curves for the \mathcal{C}_U^L and $\mathcal{M}_{\mu,t}$ with relative magnification and time delay are added for comparison. Using the time delay only leads to a slight loss in search performance. Still, the picture remains relatively close, making the identification of lensing easier. On the other hand, the model with an entirely different density profile for the lens (\mathcal{H}_t) has a significantly poorer performance, doing even worse than without the inclusion of a lens model.

given in Fig. 10.6. The FAPs for the $\mathcal{M}_{\mu,t}$ and $\mathcal{W}_{\mu,t}$ models with only the time delay included are 0.19% and 0.21% respectively. This is higher than the values for the same models with the relative magnification included. On the other hand, we see that the picture is much worse when including the wrong density profile with a very different shape in probability. Indeed, the curve found in the ROC plot from Fig. 10.5 is not comparable to the one from the other models. It is worse than for the case without a model as the curve for the \mathcal{H}_t -based model is below the one without a model for nearly all FPPs (except the highest values, where the curves are comparable). This can also be seen in the FAP, where, for \mathcal{C}_H , FAP = 1.1%, which is higher than for the \mathcal{C}_U^L statistic.

A closer comparison between \mathcal{C}_U^L and \mathcal{C}_H is given in Fig. 10.7, where we represent the values for one statistic compared to the values for another one. The two are not entirely correlated for higher values and one has more unlensed pairs with a high \mathcal{C}_H compared to \mathcal{C}_U^L . So, it is harder to distinguish between lensed and unlensed events than without a model. In the end, this means that including an erroneous lens profile can harm lensed event identification. However, if the lensed event present in the data is a “golden” lensed event (with a very high \mathcal{C}_U^L and a relatively short time delay), the wrong statistic will still enable one to detect such an event. In this case, the detection is likely to be less confident than using the correct statistic.

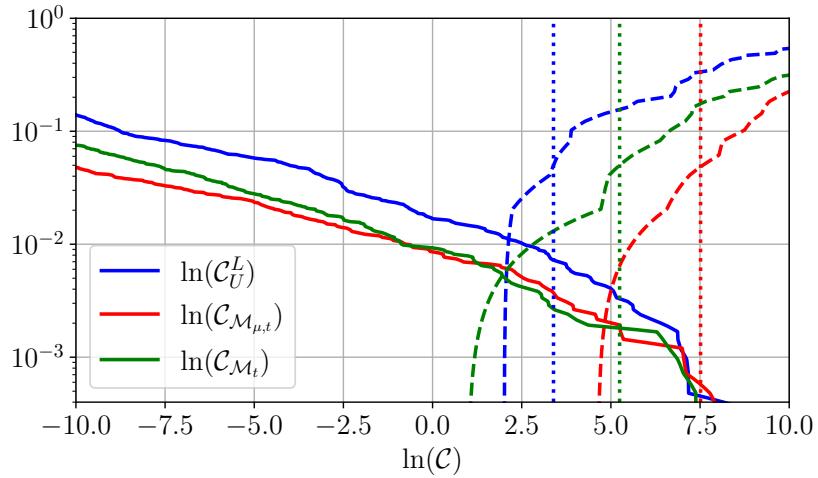


Figure 10.6: Comparison of the CDF and the CCDF for the background and foreground for the $\mathcal{M}_{\mu,t}$ model with and without the relative magnification included. The dotted lines represent the 5th percentiles of the statistics for the lensed foreground. We also represent the \mathcal{C}_U^L distributions for comparison. We see that the non-inclusion of the relative magnification leads to a larger confusion region. However, it still performs much better than without a lens model.

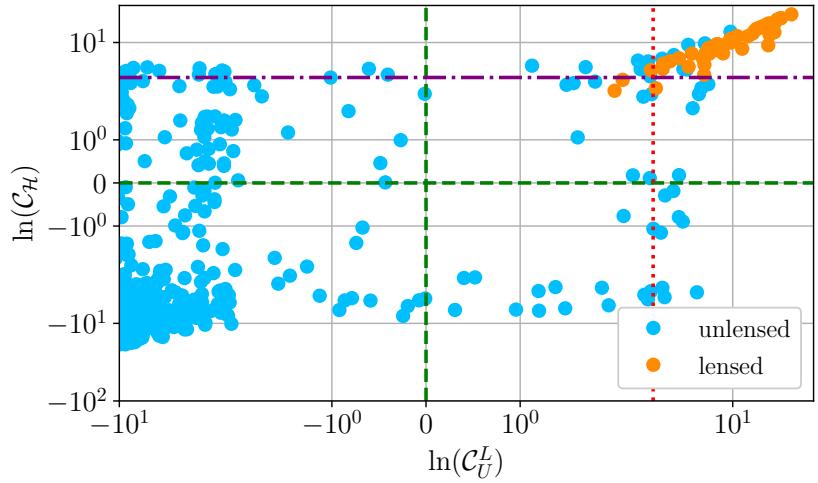


Figure 10.7: Comparison of the detection statistics for the \mathcal{H}_t model and no model at all (\mathcal{C}_U^L), with the 5th percentile (purple dash-dotted line for $\ln(\mathcal{C}_H)$, and red dotted line for $\ln(\mathcal{C}_U^L)$). One sees more unlensed events with more significant statistics for the wrong model. In addition, more unlensed events cross the 5th percentile threshold, leading to a higher FAP.

10.3.5 Analyses Using the Toy Models

In this section, we analyze the results obtained when including the toy models. These are also important as they represent some hypothetical scenarios of interest and illustrate what can happen if there are major errors in the model (for example, using an entirely biased model).

10.3.5.1 Effect of Important Errors in the Model

Here, we look at the results for models A and B, where one or two of the lensing parameters are strongly biased to higher values. Such scenarios could be observed for some galaxy-cluster lenses [e.g., 182–186]. The statistics for models A and B are written $\mathcal{C}_{\text{ModA}}$ and $\mathcal{C}_{\text{ModB}}$, respectively.

A comparison of the CDF and CCDF for model A, model B, and the $\mathcal{C}_{\text{U}}^{\text{L}}$ is given in Fig. 10.8. The two models have worse results than when no model is used. The effect is more important than for the \mathcal{H}_t case. Indeed, here, not only the shapes of the distributions for the time delay are different, but they are also focusing on an entirely different region of the parameter space. In this case, the identification would be nearly impossible for the two models. For model B, the relative magnification has the same distribution as the underlying distribution. Still, one sees the resulting model is worse, and having one correct parameter out of two is not enough to compensate when the other is strongly biased. Notably, one sees that some of the lensed pairs get a negative $\ln(\mathcal{C}_{\text{ModA}})$ or $\ln(\mathcal{C}_{\text{ModB}})$. In such a case, lensing identification would become extremely difficult as a significant part of the unlensed events has a higher significance than some of the lensed events. For model A, we observe a FAP of 1.9%, while for model B it is 2.2%. The higher value for the second model is explained by a higher number of unlensed events pushed towards a larger value. Indeed, the unlensed events tend to have larger time delays, which are more compatible with the lensed distribution for this model.

The situation represented by these models could be the one faced when analyzing a strongly-lensed event with one type of lens in mind (for example, a galaxy cluster model) while the actual lens is something else (for example, a galaxy lens). Since one does not know the true nature of the lens beforehand, performing an entire analysis based on a single model could hinder the detection of a genuinely lensed event. In addition, in our situation, the FAP is computed when knowing the underlying true distribution. However, in reality, this is not the case. Therefore, if one is not careful and uses an entirely biased model, it would have a lot of high-significance unlensed events, which could lead to false claims. The only way to ascertain the event’s nature would be to perform a background study to verify the significance of the candidate event. So, one would have to perform an extensive injection study and use the state-of-the-art BBH

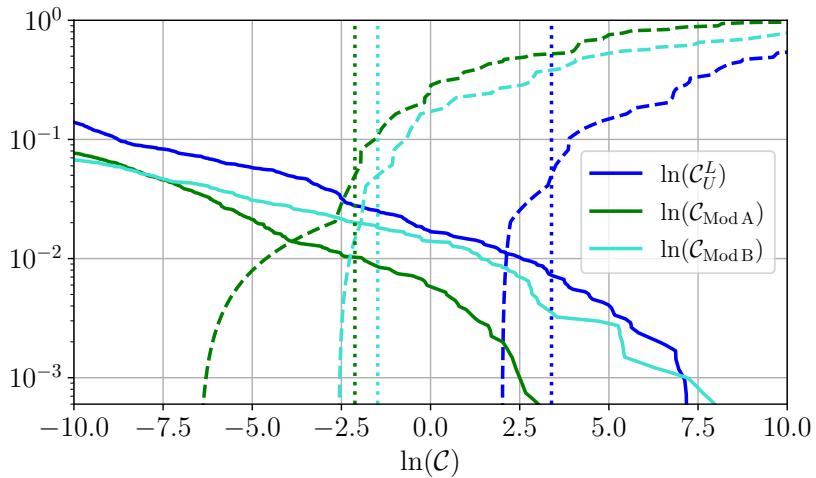


Figure 10.8: CDF and CCDF for the toy models A and B, and the coherence ratio. In this case, one sees the models are performing worse than when no model is included. This shows that including a model that would be entirely wrong compared to the true underlying distribution would make the identification of the lensed events nearly impossible. Such a case could happen when analyzing a galaxy-lensed event with a galaxy cluster lens or vice-versa.

population and lens distributions. For example, one could use the BBH population given by the LVK collaboration [46] and a lens distribution taken from a catalog, and compute the statistical significance of the candidate pair. This exercise would be analogous to the one presented in this work, except that the FAP would be represented by the number of unlensed pairs with a detection statistic higher than the lensed candidate under consideration⁸.

Finally, since the time delay distribution for galaxy cluster lensed events overlaps much more with the distribution expected for unlensed events, we expect the effect of the lensing statistics to be reduced. Hence, robust identification of a galaxy cluster lensed event would be more difficult than for a galaxy lens (see also [187] for similar results).

10.3.5.2 Effect of the Bounds on the Lensing Parameters

In this section, we focus on the alternative models C and D, which have broader and tighter bounds, respectively, than the $\mathcal{M}_{\mu,t}$ model but focus on the same region of the parameter space. This could be seen as a proxy for using the highest and lowest bounds on the model parameters. Instead of taking a hard cut on the bounds and keeping the same probability density, we rescale it to the new bounds. Therefore, in practice, we

⁸This would be one of the safest ways to ascertain the lensed nature of the observation but would also be computationally extensive, as a significant number of parameter estimation runs would be required.

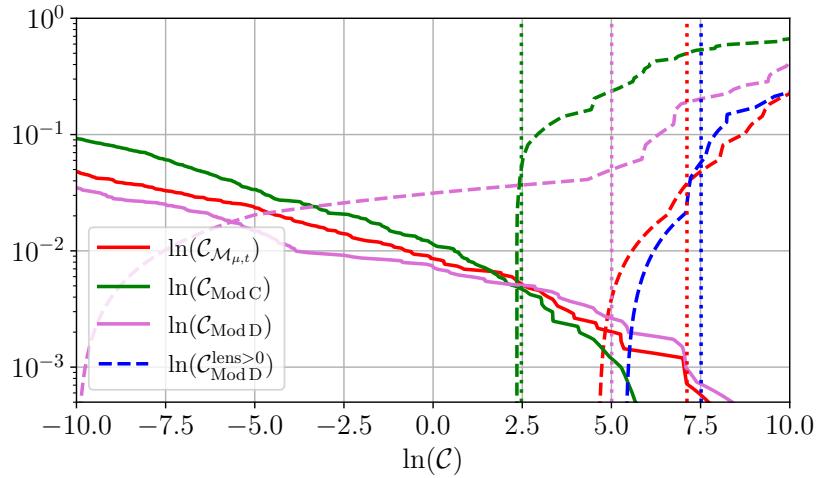


Figure 10.9: CCDF for the unlensed background and CDF for the lensed foreground for models C and D, and $\mathcal{M}_{\mu,t}$. The dotted lines represent the 5th percentiles of the statistics for the lensed foreground. The pink curve for the lensed model D foreground extends to very low values because of two lensed events with their lensing parameters outside of the bounds considered for this model. When those events are neglected (or considered as background), one gets the blue dashed curve. In this case, we see the confusion background gets much closer to the one obtained for the $\mathcal{M}_{\mu,t}$ model. Model C has larger bounds and a diluted probability density in its application domain. Therefore, we get a lower significance for the lensed events. Models C and D could represent the effect of taking the upper and lower bounds on some model parameters.

dilute the probability density for model C and condense it for model D. We denote the detection statistic $\mathcal{C}_{\text{ModC}}$ and $\mathcal{C}_{\text{ModD}}$ for models C and D, respectively.

A comparison of the performances for models C and D, and for $\mathcal{M}_{\mu,t}$ is given in Fig. 10.9. One sees that the change in bounds has consequences on the model's performance. Indeed, the two alternative models have a larger confusion background, making for a harder time differentiating lower significance lensed pairs from higher significance unlensed pairs. For model C, since the bounds on the lensing parameters are larger, more of the unlensed events have lensing parameters compatible with the lensed hypothesis. However, since the probability density is reduced, the unlensed events get less promoted, and we get a reduction of the background for the very high values. On the other hand, the lensed events get a smaller boost from the lensing statistics and therefore reach lower values. As a consequence, we also observe an increase in the FAP, with FAP = 0.47%. This value remains lower than without including any model. However, it is more difficult to confidently identify the lensed events compared to when the exact injected model is used. For model D, the FAP increases as well, since FAP = 0.64%. This is lower than without including any models, but it is still higher than the FAP obtained from using the true model. This happens mainly because of a decrease in the 5th percentile for the lensed distribution. Out of the 50 lensed pairs, 2 have lensing parameters that have values

outside of the bounds covered by model D. Therefore, they get a significant reduction in their statistic, which decreases the percentile in return. Finally, since there are still some unlensed events with compatible apparent lensing parameters, they get promoted to higher values and end up above some of the lensed events. Therefore, the background extends to values comparable to those seen for $\mathcal{M}_{\mu,t}$. If we remove the 2 events with negative $\ln(\mathcal{C}_{\text{ModD}})$, the FAP becomes 0.05%. This lower FAP is a consequence of higher values for the lensed events combined with a slight decrease in the values for the unlensed background.

This experiment shows that there is no major consequence in making errors on the bounds of the model. However, taking more stringent bounds leads to a loss in events, with some events entirely discarded. Still, if one does not account for the lensed events with a negative $\ln(\mathcal{C}_{\text{ModD}})$, the FAP for the remaining events decreases. On the other hand, using more conservative bounds helps retrieve all the events but leads to an increase in the FAP as the effect of the lensing statistic is weakened, making it less impactful.

10.4 Discussion on Triply and Quadruply Lensed Images

Typical galaxy-scale lens systems show doubly or quadruply lensed sources. For all of the doubly lensed GW signals, we need to study the detectable image pairs. However, if there exists a quadruply lensed GW source in the data, we could analyze either the brightest pair, the brightest triplets, or all four lensed images of the GW source. Needless to say, if we correctly identify 3 or all 4 images of the quadruplet, it would be a robust means of lowering the FAP. Indeed, the chances of having matching parameters between three or four unlensed events are much lower. Adding then as a constraint that the apparent lensing parameters match the theoretical distributions for a lens model makes it very unlikely for unlensed events to be mistaken for lensed ones. However, there are multiple reasons why this becomes increasingly challenging to implement.

Here, we have focused on analyzing only the brightest image pairs from the quadruplets for the following reasons. Firstly, it is computationally lighter to analyze the constraints from a pair of images rather than those from triplets or quadruplets. Secondly, the fraction of lens systems where a pair of images will be super-threshold (i.e. meet the detectability criteria) is much higher than the triplets or quadruplets for the current generation of detectors. For example, we find that about 50% of lens systems will have detectable triplets and 16% of the lenses will have all four images detectable assuming Advanced LIGO at design sensitivity. In contrast, these fractions become 70%–90% for

the third-generation detectors⁹. Thirdly, if we consider the analysis of sub-threshold lensed counterparts together with super-threshold pair of events, we are likely to introduce many uncertainties since the parameter estimation of the sub-threshold events will not be robust. It is then unclear how beneficial adding further lensed counterparts will be in improving the detection of the lens systems. Lastly, time delays are our only robustly measured observable. As a result, the correct identification of more lensed counterparts does not turn out to be as effective in practice as one might expect it to be in theory.

Analysis of triplets (and possibly quadruplets) will certainly be highly effective for 3G detectors. We leave the exercise of showing this rigorously for a future study.

In the end, having more images for a given lensed event would most likely decrease the FAP compared to event pairs, even more so when we include the lensing statistics. However, the detection of more than two super-threshold images is unlikely to happen soon, and we leave the study of the impact of lensed multiplets on the FAP for future work.

10.5 Conclusions and Outlook

In this study, we have investigated how to better distinguish between lensed and unlensed events by using a rapid joint-parameter estimation pipeline and the inclusion of a lens model in the decision process by analyzing an unlensed background with a lensed foreground. Our event pool was made to resemble as much as possible a realistic observation scenario for individual events by including changes in the PSD used to generate the noise and a variation in the number of detectors observing each event. This leads to an increase in the error made on the parameters, causing more unlensed events to be misidentified as lensed pairs by chance. On the other hand, to define our FAP as a function of the lensed population, our foreground is saturated, with an unrealistically high proportion of lensed events compared to unlensed ones. This setup enables us to probe the effect of the inclusion of lens models on the lensed and unlensed pairs under the experimental conditions of our dataset.

First, we have compared the performances of GOLUM with the results of the posterior overlap method [181], showing that comparing the strains and ascertaining the match between all the parameters decreases significantly the FAP. Based on this, we suggest a new approach for doing online searches for strong lensing. Neglecting the effect of HOMs, one could use the posterior samples from the first image obtained with traditional methods to analyze the second image under the lensed hypothesis and compare the

⁹These numbers are estimated using mock samples of [203] assuming strong lensing effects. Note that we expect demagnification of Type II images, on average, due to microlensing from the stellar population embedded within the lensing galaxies [178, 205]. This will decrease the detectable triplet or quadruplet fractions quoted here since quadruplets comprise a pair of Type-I and Type-II images each. Further accurate estimations are beyond the scope of this work.

Model	FAP
\mathcal{C}_U^L	0.75%
$\mathcal{M}_{\mu,t}$	0.07%
\mathcal{M}_t	0.19%
$\mathcal{W}_{\mu,t}$	0.095%
\mathcal{W}_t	0.21%
\mathcal{H}_t	1.1%
Model A	1.9%
Model B	2.2%
Model C	0.47%
Model D with all lensed events	0.64%
Model D without discarded lensed events	0.05%

Table 10.2: Summary of the FAP for all the models used in this work. There are two values given for Model D, one where we keep all the lensed events (including those having $\ln(\mathcal{C}_{\text{ModD}}) < 0$), and the other where we do not consider the events that would not be seen as lensed (those that have $\ln(\mathcal{C}_{\text{ModD}}) < 0$).

evidence for this image with the evidence obtained for the unlensed run. This would lead to better discrimination between the lensed and unlensed events at low latency. However, if BBHs have a larger HOM content, this method would not be entirely trustworthy, as HOMs can impact the observed parameters and lead to bias if type II images are present¹⁰.

Using our joint parameter estimation tool, we showed how to incorporate information on the relative magnification and time delay obtained from a lensed model without the need to re-do the parameter estimation, saving computational time. This can be done by reweighing the evidence obtained from the runs using the probability densities obtained from different lens catalogs. In this work, we used the results of [203], [187], and [181] to simulate three different models for galaxy lenses. We also added four toy models representing different possible observation scenarios, such as the analysis using a galaxy-cluster lens or a change in the bounds used for the model. A census of all the models used in this study and their corresponding FAP is presented in Table 10.1. We also show the FAP values obtained for all the models in Table 10.2 and represent the performances for the lensing catalog-based models in Fig. 10.10.

When comparing the discriminatory power between the lensed and the unlensed background for the different models, we have shown that, as expected, the best-case scenario is when one uses the correct model with both the relative magnification and time delay included. We have then also shown that having a slight change in the model (represented by the addition of shear in the model) leads to a slight increase in the FAP

¹⁰Posterior overlap suffers from the same caveat as it is performed on posteriors obtained during unlensed parameter estimation runs.

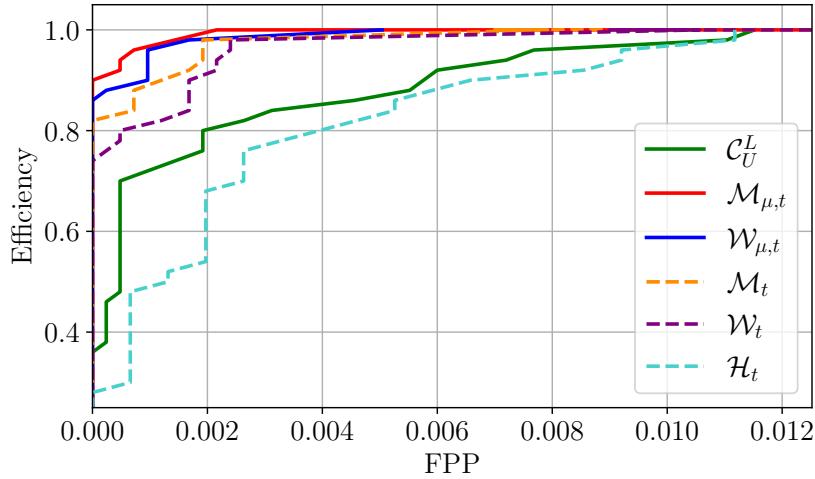


Figure 10.10: ROC curves for the existing catalogs with (solid lines) and without (dashed lines) relative magnification. The inclusion of the correct model is the best possible scenario. However, using only the time delays does not lead to a major change. Another catalog using the same density profile but using shear does not drastically change the performances either. Using an erroneous density profile degrades the performances significantly, making them worse than using no model at all.

but does not lead to drastic modifications in the identification capabilities. This implies that minor differences between the true underlying lens model and the model chosen in our analysis will not compromise the detection efficiency significantly.

We also looked at the consequences of using only the time delay in these realistic models. This leads to a slight decrease in efficiency compared to the case where relative magnification is used. Indeed, some of the unlensed events have their time delay fairly compatible with the lensing distribution but not their relative magnification. Therefore, when the latter is not included, they are less well removed from the background. Nevertheless, the increase in FAP is not important and the identification of lensing should still be easier than without using any model. We note here that the Morse factor (or phase difference) between the events has not been used in this work. However, it could also lead to more constraints and the possibility to get even better efficiency in lensing identification. The difficulty with this parameter is that the expected value is different depending on which pairs of images are seen in the lensed multiplet [203]. Hence, one would need to account for the uncertainty in the ordering of the observed images.

Next, we study the case where the time delay is obtained from the wrong density profile. Here, the identification of lensing becomes nearly impossible and the efficiency for detection is worse than in the case where no model is included at all.

This can be observed even more when looking at the results for the alternative models A and B, where we make toy models with time delays biased towards higher values. In this case, some of the lensed pairs are seen as unlensed and the confusion background is

increased compared to the scenario without a lens model. This mimics the case where galaxy-scale lenses are analyzed with a cluster-scale lens model. In such a case, the identification of lensed events would be next to impossible. This motivates the need to perform multiple lens searches in parallel, for each type of lens as the respective lens models are fairly distinct, and assuming any one model can lead to missing lenses belonging to the other type¹¹.

Finally, with models C and D, we vary the bounds of the model to understand its impact on detection. We found that the effect was a slight increase in the FAP when the bounds are widened while retaining all the events. On the other hand, when the bounds are tightened, we lose some of the lensed events. When the lensing parameters are closer to the edge of the distribution, some lensed events are discarded because they get a very low probability in the lensed hypothesis. If we compute the FAP when keeping these events in the lensed foreground, it increases because of the very low values of their statistics. However, by removing them from the lensed pool (which is what would be done in reality), we get a further decrease in the FAP. This indicates that the choice made for the bounds on the model will correspond to a trade-off between the significance and the efficiency of the detection.

In conclusion, although we know that strong lensing of GW could be detected in the coming years, identifying strongly-lensed GWs robustly is a real challenge. A large number of unlensed events leads to a significant background that can increase the false alarm risk. Still, there is hope. Even though we cannot guarantee the detection of one randomly selected lensed event in a pool of 1000 unlensed ones, the inclusion of lensing statistics in the detection process decreases the FAP, making for a higher chance to detect lensing. However, using a lens model does not guarantee the detection of all lensed events since the efficiency of the detection is sensitive to the choice of the lens model. Therefore, our suggested approach, based on this work, is to analyze first the events without a lens model using a fast joint-parameter estimation tool, and then do a follow-up analysis for the high \mathcal{C}_U^L pairs using different plausible lens models for different types of lenses, not only limited to the most likely types of lenses. This would also require the development of new lens catalogs to have statistics for other lens types than galaxy lenses (More et al. 2023, in prep.). Setting up such a framework and using extended backgrounds to find the significance of the observed events should help us identify strongly-lensed events more confidently.

¹¹We also expect this to be true if one analyzes galaxy cluster lensed events with a galaxy lens model.

Part III

MACHINE LEARNING: A TOOL FOR
GRAVITATIONAL-WAVE
DATA ANALYSIS

MULTI-MESSENGER ASTRONOMY AND THE NEED FOR FASTER ANALYSES

On the 17th of August 2017, LIGO and Virgo made the first observation of a BNS merger: GW170817 [37]. Contrary to BHs, NSs are made of matter and do not have a horizon. The interaction of this matter during the merger can lead to potentially observable EM and neutrino emissions. For GW170817, EM-counterpart searches were performed successfully. The Fermi Gamma-Ray-Burst Monitor (Fermi-GBM) [274] and the INTEGRAL satellite [275] detected the associated γ -ray signal 1.7 s after the coalescence. This event provided the first direct evidence of a link between these mergers and short GRBs. Later, other EM bands (X-ray, radio, visible, ...) were observed, making the reconstruction of the history related to the associated kilonova possible [37, 38]. This first observation was a success for GW science. The combined detection of multiple messengers allows us to improve our understanding of complex astrophysical phenomena, such as the r- and s-processes at the origin of heavier elements in the Universe. It can also lead to a better measurement of the Hubble constant and novel tests of GR, such as a measurement of the speed of GWs [276–281].

Since then, new BNS events have been detected, none with an electromagnetic counterpart [40, 282, 283]. Several explanations are possible. For example, the events were further away and the EM counterparts were too faint to be observable or the line of sight was off-axis compared to the jet emission. Observing more events with an associated EM counterpart would be useful to further improve our understanding of such events and their associated processes.

The prospects to observe multi-messenger BNSs are studied and predicted, with 0.01 to several observations a year expected in upcoming 2G observation runs [284, 285]. For 3G detectors, the CBC detection rate is larger, with tens of thousands of BNS signals detected each year. Therefore, more detection opportunities are possible [286]. However, it also comes with more data to analyze, requiring fast and accurate methods.

In the future, to maximize our chances of promptly finding EM counterparts, dedicated low-latency strategies are required. Besides, observing the EM counterparts before or closer to the merger time would help us access other stages of the binary evolution in the EM bands, and better understand the behavior of matter in these never-observed stages. Therefore, efforts are made to detect BNS events in advance [287, 288]. This would leave time for the EM facilities to prepare themselves to follow up on such events and search its sky location as soon as a sky map is issued. To make this happen, fast analysis tools are needed.

In parallel, in the last years, ML has been an emerging tool in GW data analysis (see [62] for a review). Often, it is appreciated for its speed after training, making it appealing for low-latency tasks where one can spend training time before data taking and run swiftly during the observation period. As a consequence, it is an interesting avenue to tackle BNS low-latency searches [289–291]. More details about such strategies are given in this part of the thesis.

In addition to triggering alerts about incoming BNSs, it is important to have a corresponding reliable sky map as soon as possible. Generally, low-latency sky maps – describing the probability distribution of the event in the sky – are generated using BAYESTAR [292], a rapid Bayesian framework. However, the sky localization found for a given event can change over time due to different noise characterizations, and the initial sky map is often different from the more accurate one found using complete Bayesian parameter estimation processes. Unfortunately, they arrive later as they take time and require cleaned data. Therefore, one can think of speeding up the inference methods, which can be done using ML [293–299], by using normalizing flows (explained in Sec. 6.3) for example. However, these methods still use cleaned data and infer all the CBC parameters. Also, most of them still focus on BBHs since the training is harder for BNS systems. Additionally, they may suffer some instabilities. In low latency, one is only interested in the event type and its sky location. Therefore, it can be of interest to set up a different framework where we only infer the masses and sky location. Such information can be computed rapidly using ML combined with importance sampling [300].

In this part of the thesis, we show several works done in the spirit of developing early alert and fast inference for GW signals using ML.

12

CONVOLUTIONAL NEURAL NETWORKS FOR THE DETECTION OF THE EARLY INSPIRAL OF A GRAVITATIONAL-WAVE SIGNAL

A key element in MMA is the delay between the detection and identification of a GW and its source. It is desirable to observe the event as close as possible to its merger, if not before, to not miss any part of the signal. The ability to do this depends on the GW analysis pipeline and the time needed for a telescope to point toward a possible target. For example, the Swift observatory takes a few seconds (15 s for its Burst Alert Telescope) to focus on a given sky position [301]. So, even if one wants to detect the EM emission at the merger, one needs to know where to look before it happens. Depending on the observed EM band, the observation of the merger itself would bring new insights into different aspects. In the optical and UV, it would bring more information about the r-process forming the heavy elements in the Universe [302], and about the ejecta [303]. The prompt X-ray emission would help identify the final state of the remnant [304–306], and the radio emission could reveal information on the magnetospheric interactions pre-merger [307]. The latter emission would also help testing models proposing BNS as precursors for fast radio burst [308–310].

Thus, the goal is to detect the signal during its inspiral. However, its evolution is slow, as will be detailed later in this chapter, and its amplitude is small. Hence, SNR accumulation takes time. In addition, the very early inspiral cannot even be detected as it is under the detector sensitivity, buried in noise, reducing the delay between the moment the signal enters the sensitivity band and when it merges. In the future, the delay should increase as the detectors get more sensitive, especially for 3G detectors, where the lower frequency cut-off could be as low as 1 Hz [31]. The improved sensitivity should also make for more detections and higher SNRs in the early parts of the signal.

Usual low-latency searches perform matched filtering on the full signals. Here, we want to consider the GW event before its merger, requiring an adapted pipeline. Such a setup was used in [287, 288], where the authors show promising detection capabilities in the future, with some detections done about a minute before the merger. Additionally, they also show how one can adapt BAYESTAR [292] to have sky maps before the merger. However, there is a loss in sensitivity due to the low signal SNRs.

An avenue able to circumvent some of these issues is ML and in particular CNNs.

Most of the computational cost is spent on the training part, significantly reducing the runtime when data is analyzed. Additionally, several studies have shown ML's capacities for GW detection in low-latency, obtaining sensitivities similar to matched filtering techniques [311–316]. Moreover, the inclusion of spin in ML searches seems nearly automatic as CNNs trained on non-spinning BBHs can get high performances on precessing systems [311, 312]. Some of these techniques rely on the image recognition capacities offered by CNNs and pass Q-transforms – a representation of the signal power in time-frequency bins – [317] to the network. However, generating them takes time and leads to an additional latency in the analysis.

In Ref. [289] and this chapter, we present a CNN-based framework developed to trigger when seeing an inspiraling BNS signal. We show that with a small network, it is possible to detect signals more than a minute in advance, provided the signal is loud enough.

12.1 Setup of the Study

12.1.1 Quantifying the Signal Present

The SNR (see Sec. 4.1.1 for a more extended discussion) obtained when matching a template h with data d takes the form [132]

$$\rho = \left(4 \mathbb{R} \left(\int_{f_{\min}}^{f_{\max}} \frac{\tilde{d}(f) \tilde{h}^*(f)}{S_n(f)} df \right) \right)^{1/2}, \quad (12.1)$$

where the tilde represents the Fourier domain, and $*$ is the complex conjugate. $S_n(f)$ is the noise PSD. Typically, when considering the full frequency content, f_{\min} is the minimum detector sensitivity, and f_{\max} is the maximum in-band frequency, typically set as the Nyquist frequency.

The SNR represents how well a typical template h matches the data d , made of a GW signal and noise. A matched-filtering-based search finds the template that maximizes the SNR and has an optimal performance for Gaussian stationary noise and an exactly known signal. For this type of noise, when there is no GW in the data, the SNR fluctuates around a mean value. If a GW enters the detector, the SNR increases, and when it exceeds a predefined SNR threshold, a candidate trigger event is recorded.

However, the noise from the detectors is neither Gaussian nor stationary, making the search more complex. For example, glitches can occur and lead to an SNR peak, mimicking a GW trigger. To avoid noise-triggered detections, the matched filtering-based pipelines often require the detection to be in coincidence in different detectors. Additionally, more elaborate tests downranking noise artifacts in the final candidate lists

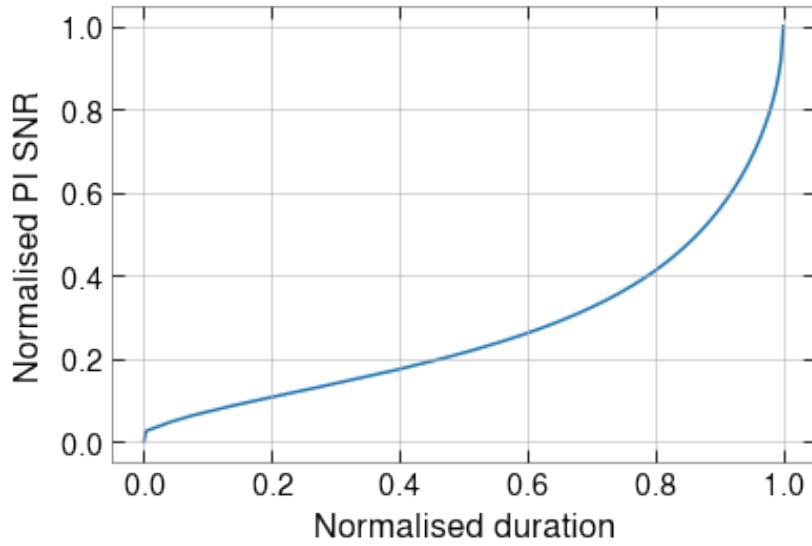


Figure 12.1: Evolution of the PI SNR as a function of the duration of the inspiral for a BNS with $1 M_{\odot}$ component masses. On the vertical axis, the PI SNR is normalized by the optimal SNR. On the horizontal axes, the duration of the early inspiral is normalized by the duration of the full template. The PI SNR increases more slowly at the start before having a rapid increase in the last 20% of the inspiral.

exist, such as the χ^2 -test [133], for example. The confidence one has about the detection is also often translated by a false-alarm rate (FAR) that gives the frequency at which noise fluctuations lead to the same ranking statistic value [318].

The optimal SNR is obtained when the template is matched with itself [132]:

$$\rho_{opt} = \left(4 \mathbb{R} \left(\int_{f_{\min}}^{f_{\max}} \frac{|\tilde{h}(f)|^2}{\mathcal{P}(f)} df \right) \right)^{1/2}. \quad (12.2)$$

In the context of pre-merger analysis, only a part of the inspiral is considered, and the SNR does not represent the signal's loudness anymore. Instead, we define it using the *partial inspiral SNR* (PI SNR), which has the same definition as the optimal SNR in Eq. (12.2), but where the template is now the partial template containing only the early partial inspiral (h_{PI}). In the frequency domain, it is equivalent, for a given waveform, to replacing the f_{\max} in Eq.(12.2) by the maximum frequency reached by the template in the part of the inspiral considered. Typically, this frequency is below 50 Hz (instead of thousands usually), significantly reducing the value of the integral.

The SNR increases more rapidly around the late inspiral and the merger than during the early inspiral. Fig. 12.1 represents the value of the PI SNR as a function of the fraction of the signal accounted for. It starts with a slower evolution, which speeds up as we get closer to the merger.

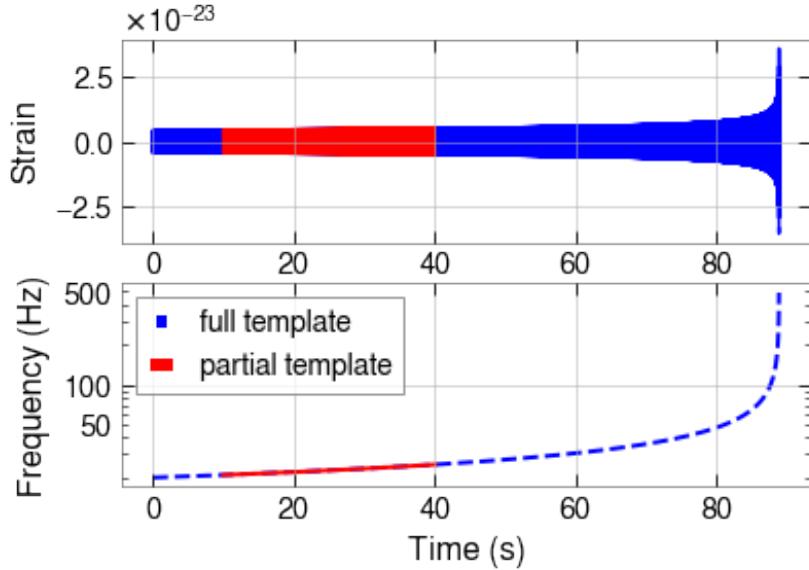


Figure 12.2: *Top:* BNS signal template with component masses $m_1 = m_2 = 2 M_\odot$. *Bottom:* Frequency evolution of the template with time. In both plots, the inspiral part considered for our ML-based approach is colored in red, while the blue signal represents the entire signal, from when it enters the sensitivity band to its merger.

The PI SNR behavior comes from the relation between frequency and time. At the lowest order in velocity¹, one finds [24]:

$$f(t) = \frac{1}{\pi} \left(\frac{G \mathcal{M}_c}{c^3} \right)^{-5/8} \left(\frac{5}{251} \frac{1}{(t_m - t)} \right)^{3/8}, \quad (12.3)$$

where $f(t)$ is the frequency at time t , \mathcal{M}_c is the chirp mass (2.50), and t_m is the time of the merger. This behavior is illustrated in Fig. 12.2, which shows the full and partial templates and their frequency evolution. Looking at the frequency evolution, when $(t_m - t)$ decreases, *i.e* when getting close to the merger time, we have a steeper increase in frequency. This is directly observable in the bottom panel of Fig. 12.2.

12.1.2 Defining Different BNS Categories

The duration of the observable CBC signal depends mainly on the chirp mass. Indeed, at the lowest order in velocity, the duration of the signal is given by [319]

¹In the early inspiral, the strong-field effects are small, which means that the expression derived for the lowest order approximates well the behavior of the binary systems.

$$\tau \simeq \frac{3}{(\mathcal{M}_c/M_\odot)^{\frac{5}{3}}} \left[\left(\frac{100 \text{ Hz}}{f_{\text{low}}} \right)^{\frac{8}{3}} - \left(\frac{100 \text{ Hz}}{f_{\text{high}}} \right)^{\frac{8}{3}} \right] \text{ s}, \quad (12.4)$$

where f_{low} is the lowest frequency in the detector sensitivity band, and f_{high} is the highest frequency reached by the binary (approximated by f_{ISCO} (2.63) for the full inspiral). From this expression, it is clear that for a fixed lowest frequency f_{low} , if the chirp mass increases, the duration of the detectable signal shortens.

Furthermore, at the lowest order in velocity, the SNR also has a simple expression [320]:

$$\rho \simeq \frac{1}{2} \sqrt{\frac{5}{6}} \frac{1}{\pi^{\frac{2}{3}}} \frac{c}{D_L} \sqrt{I} g(\theta, \phi, \psi, \iota) \left(\frac{G \mathcal{M}_c}{c^3} \right)^{\frac{5}{6}}. \quad (12.5)$$

In this expression, c is the speed of light, D_L is the luminosity distance, G is the gravitational constant, I is the frequency integral

$$I = \int_{f_{\text{min}}}^{f_{\text{max}}} \frac{(f')^{-7/3}}{S_n(f')} df', \quad (12.6)$$

and $g(\theta, \phi, \psi, \iota)$ is a function that depends on the orientation of the orbital plane and the sky position through the antenna pattern of the detectors [64]. From Eqs. (12.4) and (12.5), fixing all other parameters, one sees that if the chirp mass decreases, the optimal SNR of the signal decreases while its duration increases.

As we can observe in Fig. 12.1, the PI SNR depends on the fraction of the signal considered and on the highest frequency reached within the observation time. Therefore, observing the signal for a longer time would lead to a higher PI SNR, making its detection easier. However, we also want to detect the signal as early as possible to have an efficient pre-merger alert system. This leads to a trade-off between a high PI SNR and prompt detection in our method.

Since we know that the time evolution of the signals' amplitude is different depending on the masses, we split the BNS set into three different categories: light, intermediate and heavy BNS. For each of these, we use a separate OTW, meaning we train the networks on data with different durations. Hence, our algorithm consists of 3 CNNs, one for each category and input size. Note that the OTW is a hyper-parameter tuned in later work.

Table 12.1 summarizes the characteristics of the different categories, classified according to their chirp mass. To give an intuition for the objects' masses in each category, we present the highest and lowest chirp masses and the component masses for an equal-mass system². For each category, in addition to the constraint on the chirp mass, we also restricted the individual component masses to be between $1 M_\odot$ and $3 M_\odot$, which corresponds to a broad mass range for NSs. Note also that spin effects are absent at this order in $\frac{v}{c}$, so we considered only non-spinning BNS.

²Non-equal mass systems are also considered during the training and testing of our networks.

BNS	light	intermediate	heavy
$\mathcal{M}_c (M_\odot)$	1.13 - 1.56	1.56 - 2.09	2.09 - 2.61
f_{low} (Hz)	20	20	20
Duration (s)	100 - 180	65 - 100	45 - 65
OTW (s)	80	50	30
Fraction of signal	0.44 - 0.8	0.5 - 0.77	0.46 - 0.66
Time before merger (s)	20 - 100	15 - 50	15 - 35

Table 12.1: Summary of the BNS categories for the different CNNs. A different OTW is considered for each category because the duration of the signal changes with its component masses. “Fraction of signal” corresponds to the time the signal passes in the frame compared to the total duration. The time before merger is the duration between the end of the OTW and the end of the signal. For each OTW, the minimum and maximum component masses are restricted between $1 M_\odot$ and $3 M_\odot$.

12.1.3 Data Generation

The inputs of the NNs are 1-dimensional whitened time series made of Gaussian noise generated from the design sensitivity PSD of Advanced LIGO (aLIGO) with a GW added in some cases. The network is trained as a classifier between an *event class* (noise + template) and a *noise class* (only noise). The GW data analysis and generation have been performed with the PyCBC package [321].

We start by generating 120 seconds of colored Gaussian noise. Then, a non-spinning BNS waveform is injected. The approximant used is *SpinTaylorT4* [322], and it is generated with a minimum frequency of 20 Hz. We always generate the frame so that the signal starts at beginning of the frame, hence the time at which the signal reaches 20 Hz corresponds to the first data point in the frame.

By default, we employ the optimal sky localization and consider only the plus polarisation aligned with the interferometer’s arms to generate the various data sets. Note, however, that when we test the performance of the networks with realistic BNS populations later in this study, the sky location is not the optimal one anymore.

Since our objective is to train the networks on the early inspiral part of the waveforms, we select the desired OTW and compute the PI SNR for the generated strain. In Fig. 12.3, we plot the waveform embedded in Gaussian noise. The vertical red lines represent the portion of the strain in the OTW. The latter always starts at the beginning of the signal, when it has a 20 Hz frequency. Finally, we whiten the stretch of data under consideration and normalize its amplitude by dividing all the points by the maximum amplitude in absolute value. Therefore, all data points are in $[-1, 1]$. For the frames containing a GW, the event characteristics, such as the distance, are chosen so that the PI SNR distribution covers a wide range; see Fig. 12.5 for the PI SNR distributions.

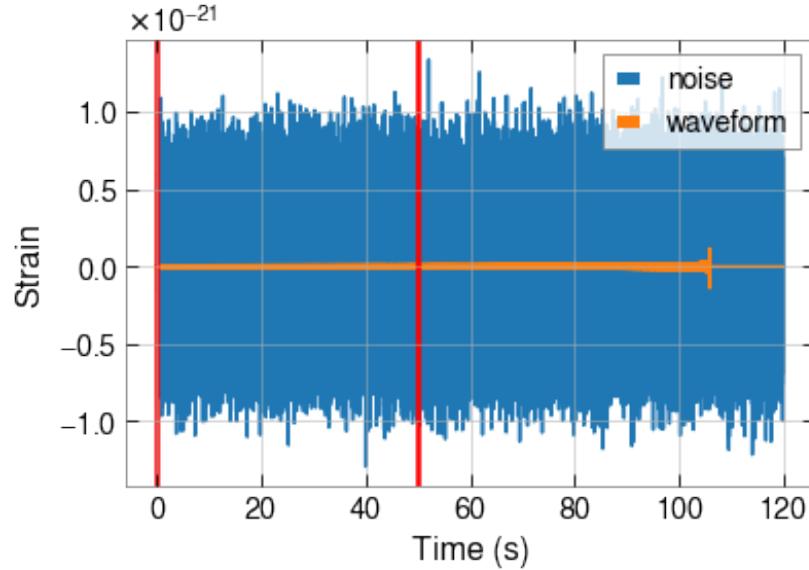


Figure 12.3: Representation of the noise and the injected waveform before the whitening. The CBC signal corresponds to a BNS with component masses of $1.8 M_{\odot}$, and placed at a luminosity distance of 100 Mpc. When training and testing the CNNs, we do not pass this full frame to the network, but only the first 50 s (denoted by the two red lines), which is the chosen OTW length for this BNS category.

12.2 Methodology

12.2.1 Convolutional Neural Network Architecture

Our goal is to perform a binary classification task, distinguishing the OTWs with GW signals from those without, with a short CNN, similarly to [311–313]. The CNNs are implemented with the *PyTorch* package [323]. We use cross entropy as the loss function and AdaMax, a variant of Adam, based on the infinity norm [229], as the optimizer. Several hyper-parameters, such as the learning rate, the batch size, the number of layers, and the kernel size, were tested. Here, we only describe the ones that provided the best performance.

After several trials, we found the best performance with the minimal computational cost for a CNN composed of 5 convolutional layers. It is found that a bottleneck structure, i.e. starting with a large kernel size, making it smaller in the middle, and enlarging it again afterward, yielded the best results. We represent the best-performing architecture

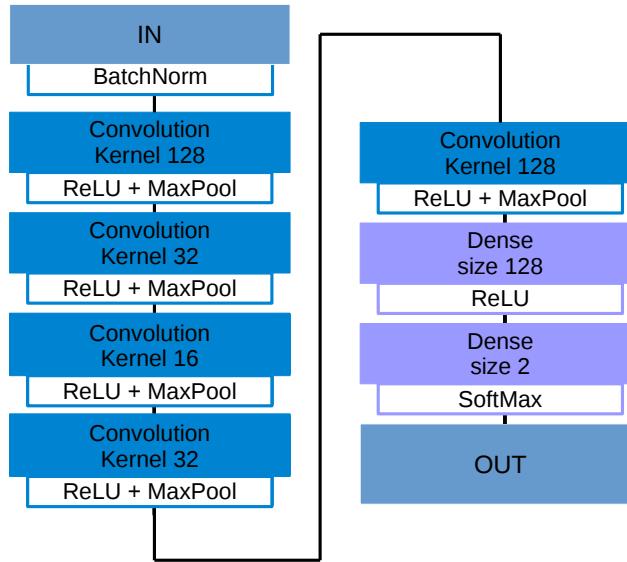


Figure 12.4: Architecture of the best performing CNN for all the categories. The input size is adapted from one BNS type to the other. The network has a bottleneck structure, with smaller convolution kernel sizes in the middle compared to their size at the start and end.

in Fig. 12.4 and give its full details in Table 12.2. The training batch size was 40 for networks 1 and 2, used for low and intermediate-mass BNSs. Due to memory issues, it is reduced to 30 for network 3, used for heavy BNSs. For the optimizer, we use a learning rate of 8×10^{-5} and a weight decay of 10^{-5} .

The network's output is the probability of the observed data belonging to the event or the noise class, depending on whether an inspiral is present in the data. The classification task is performed according to a predefined threshold, depending on the FAP.

The three networks are trained on a dataset containing 8000 frames. The data sets are balanced so that half corresponds to noise only and the other half also contains an inspiral. The testing set is balanced and made of 4000 frames.

12.2.2 Training and Testing our Neural Networks

For each category, we have a predefined OTW, given in Table 12.1. Due to the varying size of the input, we perform a binary classification task with a tuned replication of the CNN for each BNS category. We employ 80% of the training data set for training and 20% for validation. The network's performances for the training and validation sets are compared to avoid overfitting. Finally, we test each network using a balanced testing set with 4000 data frames.

Layers	Input	Output	Kernel size	Stride	Padding	Dilation	Activation
BatchNorm	1	1	-	-	-	-	-
Conv1D	1	32	128	1	0	1	ReLU
MaxPool1D	32	32	4	4	0	1	-
Conv1D	32	64	32	1	0	1	ReLU
MaxPool1D	64	64	4	4	0	1	-
Conv1D	64	128	16	1	0	1	ReLU
MaxPool1D	128	128	4	4	0	1	-
Conv1D	128	256	32	1	0	1	ReLU
MaxPool1D	256	256	4	4	0	1	-
Conv1D	256	612	128	1	0	1	ReLU
MaxPool1D	612	612	4	4	0	1	-
Dense	X	128	-	-	-	-	ReLU
Dense	128	2	-	-	-	-	SoftMax

Table 12.2: Complete architecture of our CNN. Between the last two MaxPool1D layers, we flatten all the channels to obtain an output of dimension one and length X (X depends on the OTW).

Fig. 12.5 shows the SNR and PI SNR distributions for our training and testing sets. The main difference between the two distributions is a decrease in loudness. It is due to the removal of a part of the signal. Indeed, the SNR is in the range $\sim [20, 130]$, while the PI SNR is in the range $\sim [1, 70]$. Due to the smallness of the PI SNR, the classification task becomes more difficult.

To assess the performance of each neural network, we classify its output for a given data frame into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), according to the standard *confusion matrix* [324].

Then, we define the true-alarm probability (TAP) and the false-alarm probability (FAP) as follows:

$$TAP = \frac{TP}{TP + FN} \quad FAP = \frac{FP}{TN + FP}. \quad (12.7)$$

The TAP corresponds to the number of noise + signal frames classified as such over all the number of frames that belong to the noise + signal class, whereas the FAP represents the number of misclassified noise frames over the number of frames belonging to the noise class. The networks' performances are evaluated based on the TAP for a fixed FAP.

We present all the results for a FAP of 1%. This can be considered high compared with the current GW searches, but we stress that this work is a proof of concept and our pipeline uses only one detector. By considering coincident triggers in N_d detectors,

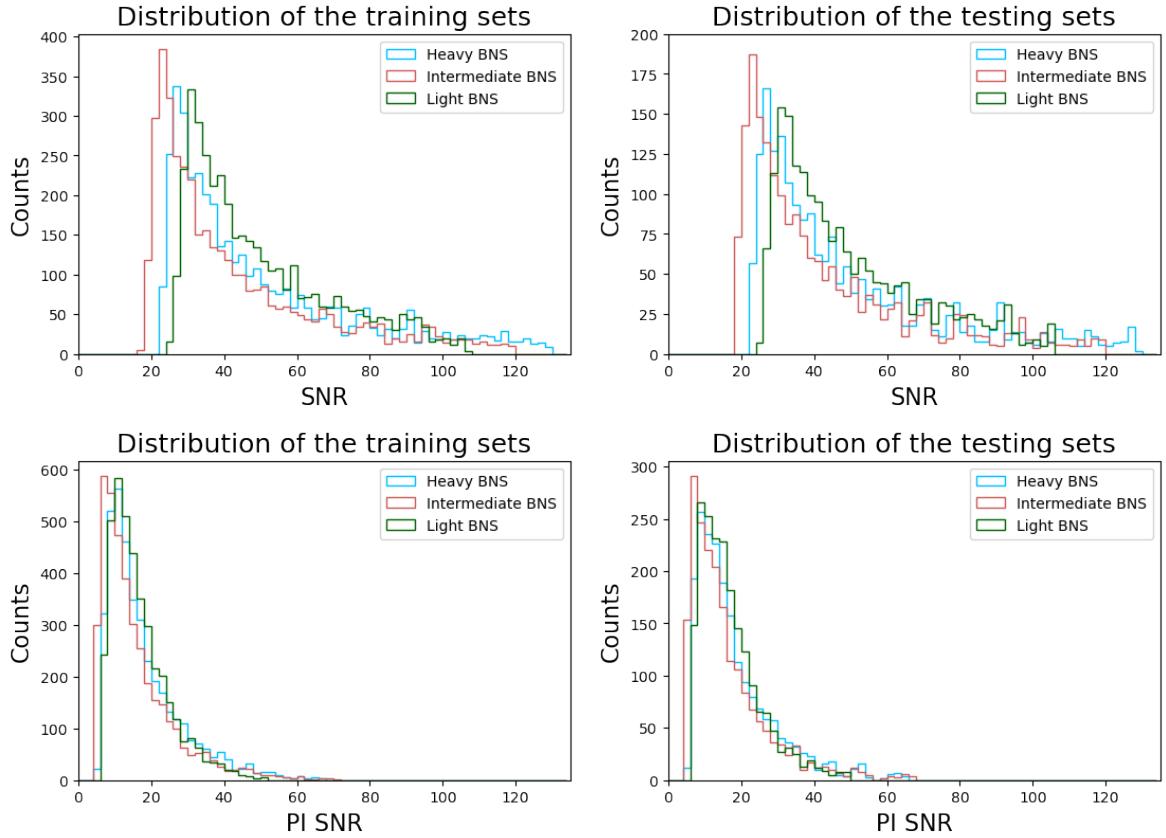


Figure 12.5: *Top:* SNR distribution for the training (left) and testing (right) sets. *Bottom:* PI SNR distribution for the training (left) and testing (right) sets.

the FAP will roughly scale as 0.01^{N_d} , naturally reducing the risk of false claims³.

12.3 Results and Discussion

In this section, we first discuss the performance of the three networks. Then, we look at our networks' performance when applied to a realistic BNS population. Finally, we discuss a first simple attempt at curriculum learning.

³This is an approximation where we assume that the three channels are independent and that, at each instant, each CNN has a 1% chance to claim a false detection.

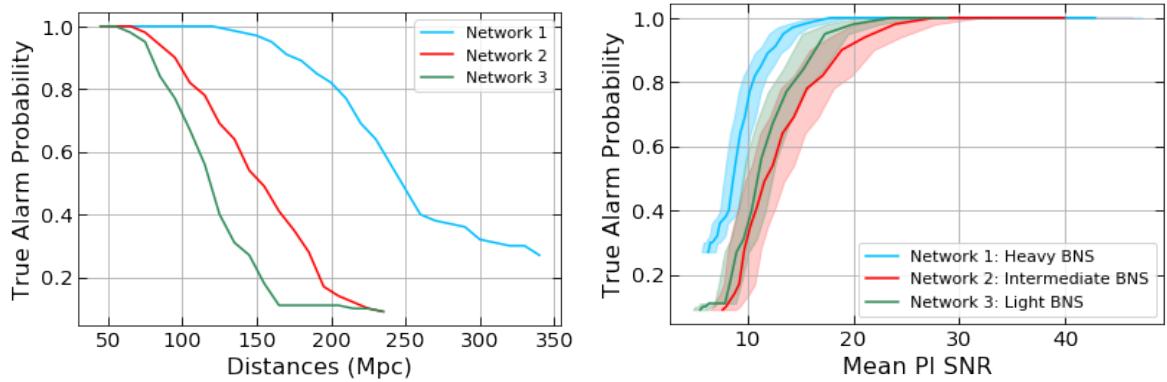


Figure 12.6: *Left:* Results of the three networks, each trained on its category, as a function of the distance. *Right:* Mean (μ) PI SNR and its standard error (ε), $\mu(\text{PI SNR}) \pm \varepsilon(\text{PI SNR})$ and a confidence of 2σ , represented by the colored band, are plotted for each network. For each figure, the FAP is fixed at 0.01.

12.3.1 Performance of the CNNs

In Fig. 12.6, we plot the TAP as a function of the distance and the PI SNR for each BNS category individually. The network trained on heavier objects can reach higher distances. From Eq. (16.3), we observe the same behavior: for smaller chirp masses, we need to decrease the luminosity distance to keep the same SNR value. We obtain the best performance for the heavy BNS category. The intermediate and low categories have very similar performance, where we see that the 2σ interval for both overlaps when considering the PI SNR. Since the architecture of the network was optimized for the heavy category, we expect it to perform best for these BNS inspirals. Adapting the networks' structures for each category could help enhance the performance of each of them.

Note that the CNNs are sensitive to the accumulation of the signal. To confirm this, we trained and tested the networks on data with low-frequency cut-offs set to higher values than the usual 20 Hz. This is a way to reduce the PI SNR of the injected signal while maintaining the same maximum amplitude. For the testing set, we obtained an 88% TAP for a cut-off at 20 Hz, and 71% for a 26 Hz lower frequency, showing that the CNN is sensitive to the PI SNR for a fixed maximum amplitude. Similarly to matched filtering, a CNN is designed to recognize patterns and, in this context, the larger PI SNR means that the signal is present for a longer time.

From Fig. 12.6, we see that network 1 can reach distances larger than 60 Mpc before its TAP departs from 100%. Since the first BNS detected (GW170817) was located at a distance of about 40 Mpc [37], our method has a high probability of detecting similar signals when present in noise at design sensitivity. So, our method can recover a realistic

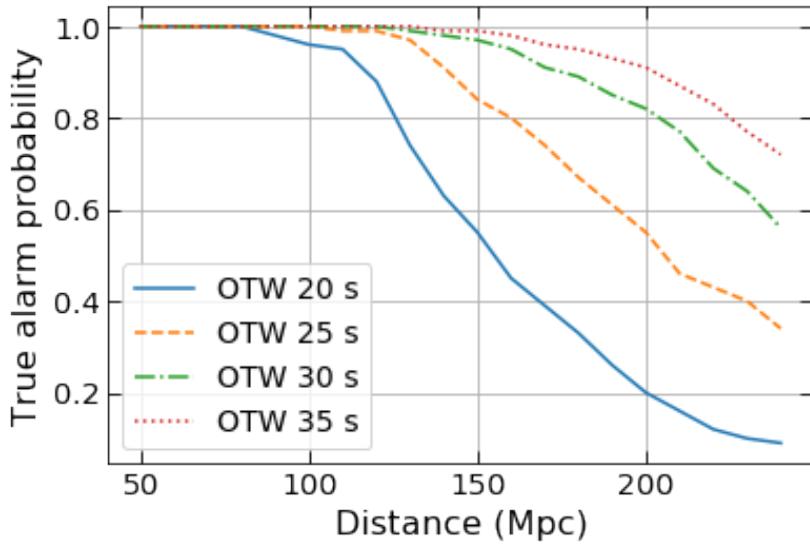


Figure 12.7: Representation of the performance of the CNN trained on the heavy BNS systems for different OTWs. A longer window gives a higher number of detections. However, it also means the detection happens closer to the merger time. The mean times before the merger are 35, 30, 25, and 20 s for the 20, 25, 30, and 35 s OTW, respectively.

signal from colored Gaussian noise when only the inspiral part is present. Network 2, which is trained on intermediate BNSs, can perform better at higher distances compared to network 1, which is expected based on the chirp mass - PI SNR relation. Finally, Network 3 has a TAP of 100%, even for a distance of 125 Mpc. Hence, the detection efficiency remains high for distances similar to that of GW190425, the second BNS discovered by the LVK collaboration [42].

We now perform a series of tests to evaluate the influence of the length of the OTW. This is an important hyper-parameter representing the fraction of the signal seen for a given event. It needs to be optimized to have as many detections as possible while keeping a long enough delay between the trigger and the merger time. In Fig. 12.7, we show the TAP for Network 3 when using different OTW. As expected, a larger OTW increases the TAP but is associated with a shorter time remaining before the merger.

We also test whether a network trained on a given category can find signals that belong to a different one. We concentrate on Network 3, trained to detect heavy BNSs and check whether it manages to detect intermediate BNSs. For this, we decrease the OTW of intermediate BNSs to 30 s, to feed the data set to Network 3. We find that the TAP decreases significantly. Network 2, trained to detect intermediate BNSs, yields a TAP of $\sim 68\%$, while Network 3 reaches only $\sim 16\%$. This is also understandable in terms of PI SNR. The reduction of the OTW duration leads to a decrease in the PI SNR, and we already established that this is a key parameter for detection.

We now compare the time needed for our CNN and matched filtering to analyze a data

frame. When applying matched filtering on a 50 s early-inspiral frame, similar to those passed to the CNNs and with only the optimal template, the computation time is ~ 0.05 s⁴. This is just the time needed to get the SNR in matched filtering when computing the statistic using the injected template. In this traditional method, several templates are tested, and the trigger is not only assigned an SNR, but also other statistics, such as the FAR. As a consequence, the time to get the final information is longer [325]. Analyzing the same frame using our CNN on a *Nividia GeForce RTX 2070 SUPER* GPU, we get the probability of an inspiral to be present in ~ 0.005 s. Therefore, the time to analyze the frame and get a probability of there being an inspiral is improved by a factor of ~ 10 .

12.3.2 Test on a Realistic BNS Population

To better grasp the performance of our networks compared to matched filtering, we also test them on a simulated realistic population of BNS systems. Therefore, we compute both the optimal and PI SNRs for each BNS with a high-frequency cut-off of 32 Hz, similar to what was done in [287]. When performing the run with a high-frequency cut-off and the test with our CNNs, we only consider the events with a matched filtering SNR higher than 8. This basic computation is performed for the high rate presented in [326].

The cut-off frequency of 32 Hz has been chosen to give results comparable to those in [287] while having in-band times corresponding to the OTWs defined in Table 12.1.

The population synthesis is performed using the code of [319], with minor changes to suit our framework. For example, the PSD employed is the same as for the noise generation, the low-frequency cut-off is 20 Hz, and we generate the equivalent of 5 years of data.

One shortcoming of this procedure to generate a realistic population of BNSs is that, although it is fast, it is based on analytical approximations. As a consequence, we do not inject the signals in noise to compute the SNR and cannot calculate the matched filtering FAR for such frames. So, we cannot use the criterion of Ref. [287] (an SNR threshold followed by a FAR), and the direct comparison is non-trivial. Our procedure confirms the difficulty to detect those events with matched filtering methods.

Once we have selected the events based on the analytical approach, we inject them in design-sensitivity noise and pass the frames to the CNNs. Fig. 12.8 represents the events detected, missed, and misidentified. We also generate the same noise for each event without injecting the BNS into it. We test our networks on these pure noise frames to highlight the FPs. As shown in Fig. 12.6, the networks detect most of the BNSs which have a sufficiently high PI SNR. We also want to emphasize that matched filtering applied for pre-merger alerts needs a PI SNR above a threshold to lead to a trigger.

⁴All the tests were done on an *Intel(R) Core(TM) i7-8650U* CPU.

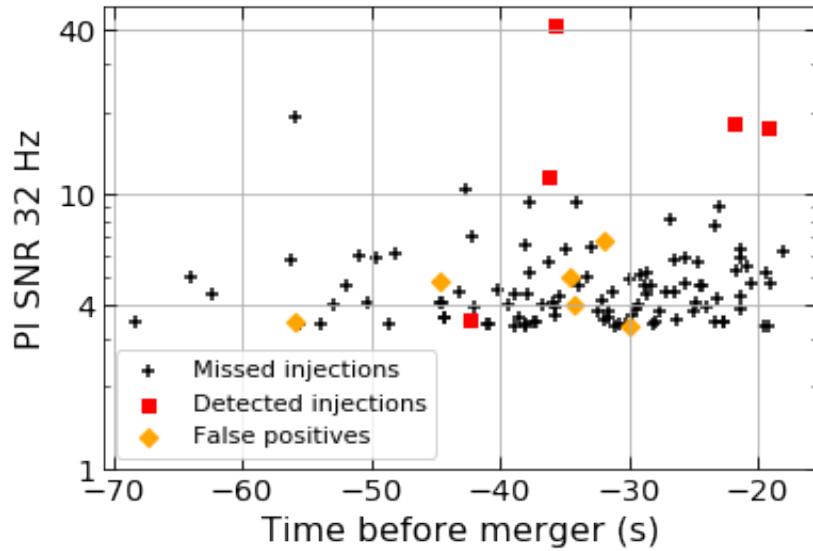


Figure 12.8: The PI SNR for a low-pass filter at 32 Hz for each BNS with a full SNR higher than 8. The black crosses represent the events missed by the CNNs, and the red squares are the events correctly found. The orange diamonds are triggers that correspond to noise fluctuations (false positives). This represents the detection one could expect in O4 data using our network. Since most events have lower SNR, we only detect the loudest events.

This threshold depends on the framework and the number of detectors included. We can see that, if one chooses an SNR threshold of 8, our results are comparable to those of matched filtering. Nevertheless, CNNs are much faster⁵.

A key feature employed in [287] is the network of detectors. Requiring coincident detections in the different interferometers helps to remove signals due to noise artifacts. Another advantage is that the data can accumulate in several detectors simultaneously. Additionally, sky localization is found using the data in the three detectors [327]. For a NN, the input will have a certain number of channels, one for each detector. Then, the input will be convolved through the network, finding relationships between the different channels. This should decrease the FAP of our detector network and enable us to perform sky localization. This will be explored in future work.

12.3.3 Basic Curriculum Learning Exploration

Aside from the architecture, another crucial factor in the development of DL algorithms is the training procedure. From the population analysis, we conclude that the networks see the loudest events, i.e. those with the highest PI SNR in the OTW (or the highest

⁵Here, we neglected the latency needed for the data transfer. During the O3 run, the latency was ~ 20 s. Eventually, one would like to bring this down to ~ 7 s [287].

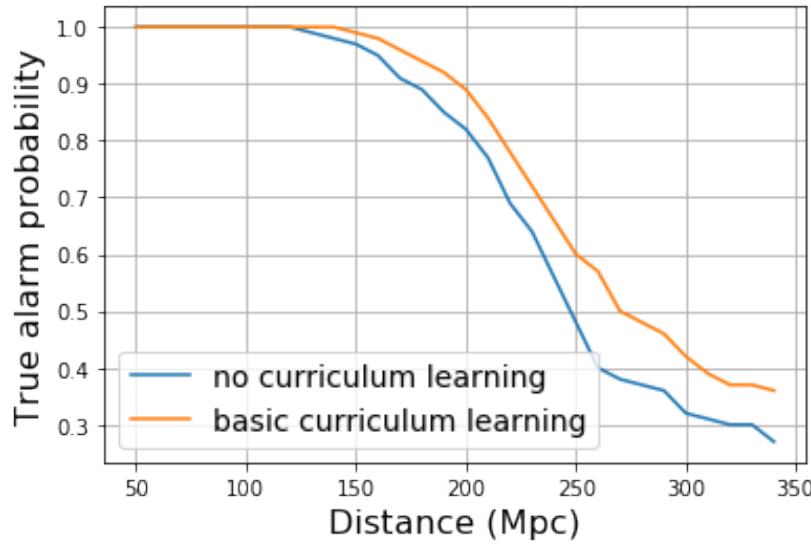


Figure 12.9: Comparison of the TAP as a function of the distance for the GW sources with and without curriculum learning for the heavy BNS class. A rudimentary curriculum learning setup helps improve the TAP at higher distances. Note that the blue curve is the same as in Fig. 12.6.

SNR in the detector for the full template). The networks have been trained on a wide distance range for the events (hence a wide PI SNR range), but it is hard for them to detect smaller PI SNRs, as we can see in Fig. 12.6. A way to overcome this obstacle is by training the CNNs with *curriculum learning* (see Sec. 6.1.5 for a more in-depth introduction). The main idea is to train the network on batches of PI SNR, first on the easy examples, namely the frames with the highest PI SNR. Then the difficulty is increased iteratively by decreasing the PI SNR until the hardest examples are considered, namely the frames with the lowest PI SNR (see [312] or [328] for an example).

With this idea in mind, we generate an extra batch of training data with higher distances and lower PI SNR. Thus, we train on the first data set, store the weights, then train on the newly generated set beginning with the weights saved previously. The results of this test can be seen in Fig. 12.9. The TAP increases significantly even if we are using only one extra data batch. As a consequence, we expect our networks' efficiency to increase substantially once they are trained using curriculum learning.

12.4 Conclusions

In this study, we introduced a new approach based on short CNNs for pre-merger alerts. We have shown that it is possible to detect BNS events when only part of the early inspiral is present in the data stretch under consideration. For this purpose, we have introduced

three different neural networks, each trained on a particular range of chirp masses for the BNS systems. Such developments are important for MMA, as the prediction stage is computationally less expensive and usually faster than traditional matched filtering. We have also shown that our method can recover signals from a realistic BNS population simulated at design sensitivity and compared our detection statistics to those obtained with matched filtering pipelines. In addition, we also suggested some improvements in the training method, as well as in the structure of our CNNs, to enhance their performance further, leading the way to a competitive pre-merger alert system.

This work presents a proof-of-concept ML-based early-alert system, and we will continue to build upon this basis to upgrade our networks and get better performances. The next steps, which will probably require more complex networks, are the consideration of multiple interferometers and computing the sky location. Furthermore, curriculum learning will systematically be deployed as this will allow us to train on a larger dataset with smaller PI SNR. Indeed, the training set currently has a minimum PI SNR of around 8. With curriculum learning, it will be possible to lower this value. A fourth CNN trained to retrieve the full BNS signal, regardless of its category, will be built. This will complete the pipeline as the events not detected based only on their inspiral would still be found in low latency. Another recent approach [290] used CNNs to remove some non-linearly coupled noises and detect the early phase of a gravitational wave signal. This further reinforces the case for using CNNs in early-alert systems.

Various works have shown that ML-based algorithms can help GW astronomy. In this work, we have shown that it can also be used to solve one of the challenges that will arise in the future, namely the early detection of BNS mergers in the context of MMA. However, we still want to improve the performance and add some features, such as sky localization. These are the next milestones, probably requiring more complex networks and more advanced training methods.

13

CONVOLUTIONAL NEURAL NETWORK FOR GRAVITATIONAL-WAVE EARLY ALERT: GOING DOWN IN FREQUENCY

Third-Generation Detectors: Interesting Science, Though Analyses In the previous work [289] and Chapter 12, we presented a proof-of-concept pipeline to perform pre-merger alerts using ML, in particular one-dimensional CNNs. Others have had a growing interest in early warning with ML [290], where they show that using a filtering algorithm to clean the data helps improve detection capabilities. The work presented in this chapter and Ref. [291] builds upon the framework developed in Chapter 12. We construct more robust techniques and apply them to more realistic scenarios. We now accommodate a detector network made of the two LIGO and Virgo detectors. Additionally, we also decrease the starting frequency of the signals and aim at detecting signals at a lower maximum frequency. This is crucial since finding signals at a lower maximum frequency means detecting them with an increased time before the merger.

The main changes compared to the previous chapter are

- Addition of spin effects in the BNS waveform ;
- Use random sky location for the events in the data ;
- Consideration of different types of noises: simulated O3 and O4 noises, and real O3 noise ;
- Decrease of the minimum frequency cut-off for the data, going from 20 Hz to between 10 and 15 Hz ;
- Longer and fixed input signal of 300 s sampled at 512 Hz, allowing for the analysis of all BNS events without the need for classes ;
- Implementation of a more robust curriculum learning scheme [329].

13.1 Method and Setup

As in the previous chapter, the signals are characterized by their SNR; see Eq. (12.1). However, since we consider a network of detectors, the signal’s total loudness is expressed via the network SNR (see Eq. (4.12)), the square root of the quadratic sum of the individual SNRs.

As detailed in Sec. 12.1.1, when considering only the early inspiral, we quantify the signal’s loudness with the PI SNR, computed similarly to the SNR, except the maximum frequency is the highest in-band frequency. This can also be combined into a network PI SNR, similar to the SNR and the network SNR.

13.1.1 Data and Training Strategies

In the previous chapter, we showed the possibility of detecting the early inspiral of a BNS injected in Gaussian noise. In this work, we want to turn to a more realistic scenario using real O3 noise. To investigate the difference in performance between Gaussian and real noise, we also inject the signals in colored Gaussian noise generated from the O3 representative PSDs [327]. In addition, to assess the performance of our network in future observation runs, we consider colored Gaussian noise generated from the predicted O4 PSDs [327, 330].

The corresponding PSDs are represented in Fig. 13.1. To generate a frame of simulated O3 Gaussian noise, we use the PSDs from [327], provided by PyCBC [321]¹. To obtain O3 data, we download the detectors’ strain directly [40, 126, 267] using the GWPy package [331]. To resemble a real-time search, these downloaded strains are the ones recorded in low latency, meaning that they are not filtered and cleaned as extensively as the final noise². To generate the O4 Gaussian noise, we use the predicted O4 PSDs [327, 330]³.

Since the problem at hand can be solved as a classification task, we need a data set containing two classes: noise and noise plus inspiral, also referred to as injections. For the latter, we generate waveforms using the *SpinTaylorT4* approximant [322]. We

¹The PSD used for Gaussian O3 LIGO is *aLIGOaLIGO140MpcT1800545*, the one for Virgo is *aLIGOAdVO3LowT1800545*, both are provided by LIGO and Virgo and implemented via PyCBC [321].

²To download the real O3 data, we use the channels H1:GDS-CALIB_STRAIN, L1:GDS-CALIB_STRAIN, V1:Hrec_hoft_16384Hz, and the frame type: H1_llhoft, L1_llhoft, V1Online in GWPy.

³The LIGO and Virgo PSDs used for O4 correspond to the ones shown in Fig. 1 of [327], with the BNS detector horizon at 160 Mpc for the LIGO detectors and the horizon at 120 Mpc for the Virgo detector.

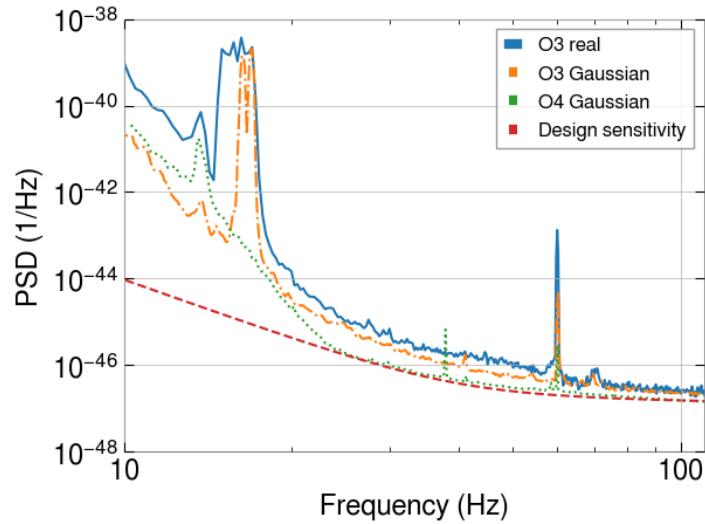


Figure 13.1: Representation of the different PSDs for the Livingston detector. They are used to generate different data sets. We also show the design sensitivity PSD provided by PyCBC [321] used in [289].

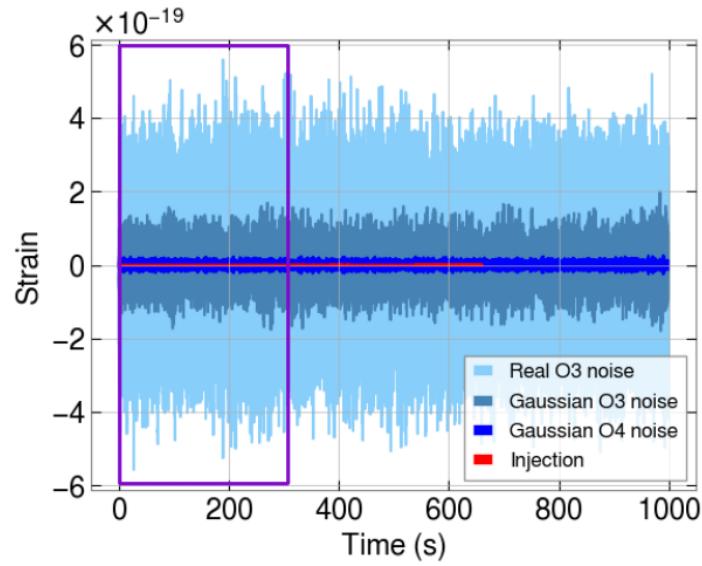


Figure 13.2: Representation of the different types of noises for the Hanford detector used together with an injection similar to GW170817, i.e. with neutron star masses of 1.46 and 1.27 M_{\odot} [37]. Note that only the part in the rectangle is passed to the network.

choose uniformly distributed component masses between 1 and $3 M_{\odot}$ to cover all the possible BNS systems [332]. The sources are distributed uniformly over the sky, and we also include the spin effects. With these parameters and a lower frequency of 10 Hz, the simulated signal is always longer than 300 seconds. In such a way, the network's inputs contain only the early inspiral part (see Fig. 13.2 for an illustration). After injecting the simulated signal into the noise, the frames are whitened, and we apply a low-pass filter at 100 Hz⁴ and a high-pass filter at 10 Hz. Afterward, the final data is renormalized, making all the values in the frame between -1 and 1 . This is the input data of the network, and we refer to a single sample as a *frame*.

For O3 real noise, some significant peaks can appear in the whitened strain due to non-Gaussian effects (see Fig. 13.3 for an illustration). In our approach, these effects are vetoed by zeroing them out as they would lead to issues in the normalizing scheme. The large peak would be the maximum value used to divide the values for all the other points, making the other points in the time series too small. This confuses the NN. Therefore, we remove these artifacts from the data. The vetoing is done according to the z-score:

$$Z_i = \frac{x_i - \mu}{\sigma} \quad (13.1)$$

where x_i is the strain value for a point i in the time series, μ and σ are the mean and the variance of the time series, respectively.

The z-score is computed for each point. We also calculate its standard deviation for the entire series. All the data points with a z-score larger than five times the standard deviation are set to zero, removing the dominant peaks from the data (see Fig. 13.3). Then, the normalization and analysis are performed on the vetoed data frame. Generally, when applying this procedure, we only veto a few points in a row.

For the training and testing, we choose a distance distribution such that the distribution in PI SNR is an inverse Gaussian with a mean of thirty-five and a shape parameter of one hundred⁵. Despite having a large data set containing one million frames, we have observed a low performance when we decrease the maximum frequency to ~ 25 Hz. This is because the CNN is sensitive to a variation in frequency, and for earlier inspiral phases, the signal is more monochromatic.

To detect events earlier, it is key to decrease the maximum frequency seen by our model. Therefore, we change the training strategy and use curriculum learning [329] on the maximal frequency seen by the network. In previous works [328], curriculum

⁴Since we focus on the early inspiral of the signal, it is not expected to reach higher frequencies.

Therefore, keeping all the data would not help in the detection since it would only contain noise.

⁵We found that the inverse Gaussian (Wald) distribution fits better our goal. Indeed, this distribution gives a few very high PI SNR events that enable the network to start its learning process. The PDF of an inverse Gaussian is

$$f(x) = \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right),$$

where μ is the mean and λ is the shape parameter.

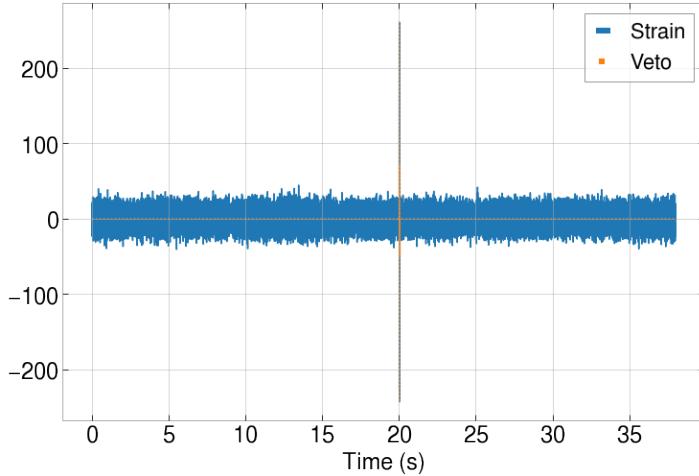


Figure 13.3: The blue curve represents O3a noise after the application of the whitening, the low-pass filter, and the high-pass filter. The orange curve shows the vetoed part.

learning applied to the SNR has already led to increased performances. Its principle is to train the network on easier data first (on data with a high maximum frequency), then gradually increase the difficulty (on data with a lower maximum frequency). The network is then iteratively trained on each set. To prevent the network from forgetting what it has learned, we keep all the data of the previous steps while adding new ones. To that effect, we generate five different training sets. The parameter distributions for the injections stay the same, except for the maximal frequency seen by the network. It is now a Gaussian distribution with a standard deviation of 2.5 Hz, and a mean depending on the data sets. More information about the data sets is presented in Table 13.1. Each training set contains 20000 frames, and half of them contain an injection. 20 % of each training set is used for validation during the training. For each step, we train for six epochs since it is enough to make the loss converge without facing over-fitting. The use of curriculum learning allows for improving the performance on data set 3, 4, and 5 with a maximum frequency of, respectively, 30, 25, and 20 Hz, while maintaining the performance at higher frequencies.

The training on the real noise data is done similarly. It is done only with noise coming from O3a, the first half of O3 [267]. We veto the periods corresponding to real events from the GWTC-2.1 catalog [267]. For all the testing, we used noise coming from O3b, the second half of O3 [40, 126]. During O3 there are times when not all detectors are online. To account for this, we fill the CNN entry corresponding to the detector with a vector of zeros when it is offline. This way, our network can perform the search regardless of the number of interferometers available.

For the training parameters, we use a batch size of 50. The learning rate is 8×10^{-5} , and the optimizer is AdaMax with a weight decay of 10^{-5} . AdaMax is a variant of Adam, based on the infinity norm [229]. In the previous chapter, we have seen that

Data set	Max. Freq	Min. Freq	Min. TBM	Max. TBM
Data set 1	40 Hz	12.9 Hz	7 s	44 s
Data set 2	35 Hz	12.8 Hz	10 s	63 s
Data set 3	30 Hz	12.6 Hz	15 s	95 s
Data set 4	25 Hz	12.3 Hz	24 s	115 s
Data set 5	20 Hz	11.7 Hz	45 s	280 s

Table 13.1: Each data set corresponds to a maximum frequency seen by the networks, which in turn leads to a minimum frequency and a Time Before Merger (TBM). The maximum and minimum frequencies presented are the mean values in each data set. The maximum and minimum TBMs are for equal mass BNSs with $1 M_{\odot}$ and $3 M_{\odot}$ NSs, respectively.

using AdaMax leads to faster convergence during training.

We use the weighted cross-entropy loss [333]. At first, we employed the cross-entropy loss, which is standard for classification problems. However, this led to a large number of false positives. To avoid this, we weight the loss [328] by a factor of 0.4 for the frames with an injection. This reduces the chances that the network classifies a noise-only frame as an event, reducing the number of FPs. We tried multiple values for the weight and found that, for the task at hand, a factor of 0.4 translates into a reduction of the number of FPs while maintaining the number of TPs. The duration of the training is about one day on a *NVIDIA Tesla V100-PCIE-16Gb* GPU.

13.1.2 Description of the Neural Network

The NN’s architecture is similar to the one in Chapter 12 (see Fig. 12.4), where we tested multiple structures to find the most effective one. We use the PYTORCH package [323] to implement the architecture. The network takes 300 s of data for each available detector. In other words, it has three input channels, each corresponding to one of the three detectors (Hanford, Livingston, and Virgo)⁶. It is composed of a batch normalization layer, followed by five blocks made of a convolution layer, a ReLU activation, and a pooling layer. For the convolution, the kernel sizes are successively 16, 8, 4, 8, and 16. For the pooling layers, the kernel size is always 4. The stride is 1 for the convolution layers and 4 for the pooling layers. After these blocks, we add two linear layers with sizes of respectively 128 and 2 interspersed by a ReLU activation. The final layer is a softmax layer returning a probability vector.

⁶The Conv1D layer as implemented in PyTorch allows us to give as input any number of channels, see <https://pytorch.org/docs/stable/generated/torch.nn.Conv1d.html> [323].

13.2 Results and Discussion

This section discusses the results of the various changes made to decrease the maximum frequency required for detection. We also compare the performances for the different noises and look at the effect of curriculum learning on the performance. Finally, we also investigate and discuss different ways to decrease the FAR.

13.2.1 Performance of the Network

The testing sets are built with the same distributions as the training sets, see Table 13.1. The other parameter distributions are the same as for the training sets. Each of the test sets contains 4400 frames, half of which are pure noise and the other half contain an injection. The total size of the test sets for a type of noise is 22000 frames.

The efficiency of our network for the different steps of curriculum learning can be seen in Fig. 13.4. We define the TAP and the FAP as in Eq. 12.7. In the top panel of Fig. 13.4, we present the results for the three lowest maximum frequency data sets. The higher maximum frequencies have performances similar to the 30 Hz data set. For the data sets with a maximum frequency > 25 Hz, an efficiency of 50% is obtained at ~ 15 PI SNR, while the efficiency reaches 100% at 30 PI SNR. This is not the case for the data set with a maximum frequency of 20 Hz, where the TAP is lower. This is expected since the detectors' sensitivity worsens at lower frequencies, typically under 20 Hz, see Fig. 13.1, left panel. In all the figures shown in this work, the FAP is fixed at 1%.

Similarly, we did the same test for the real O3 and the simulated O4 Gaussian noises. The different results are summarized in the bottom panel of Fig. 13.4, where each curve represents the results for the whole test set. In terms of PI SNR, the efficiencies for O3 Gaussian noise and O4 Gaussian noise are similar. However, since the noise floor is lower in the O4 case, the network can probe higher distances in this case. The performance for real noise is a bit worse than for the two Gaussian cases. The network needs a slightly larger PI SNR to achieve the same performance. For example, the network requires a PI SNR of 20 to have an efficiency of 50% in the case of real O3, whereas it only needs a PI SNR of 17 to reach the same sensitivity in the two other cases. Even if some glitches and non-Gaussian features are present in the data, the network still reaches a high performance, provided the PI SNR is high enough. To resemble a real-time search, the noise is downloaded from genuine low-latency strains. Therefore, it has a lower quality, explaining the reduced performance.

After testing the network on independent 300 s-long frames, we generate frames of 1000 s and inject a complete GW signal into them. Then, we slide a 300 s window over the frame, pass the data in the window to the CNN for each step and make a prediction.

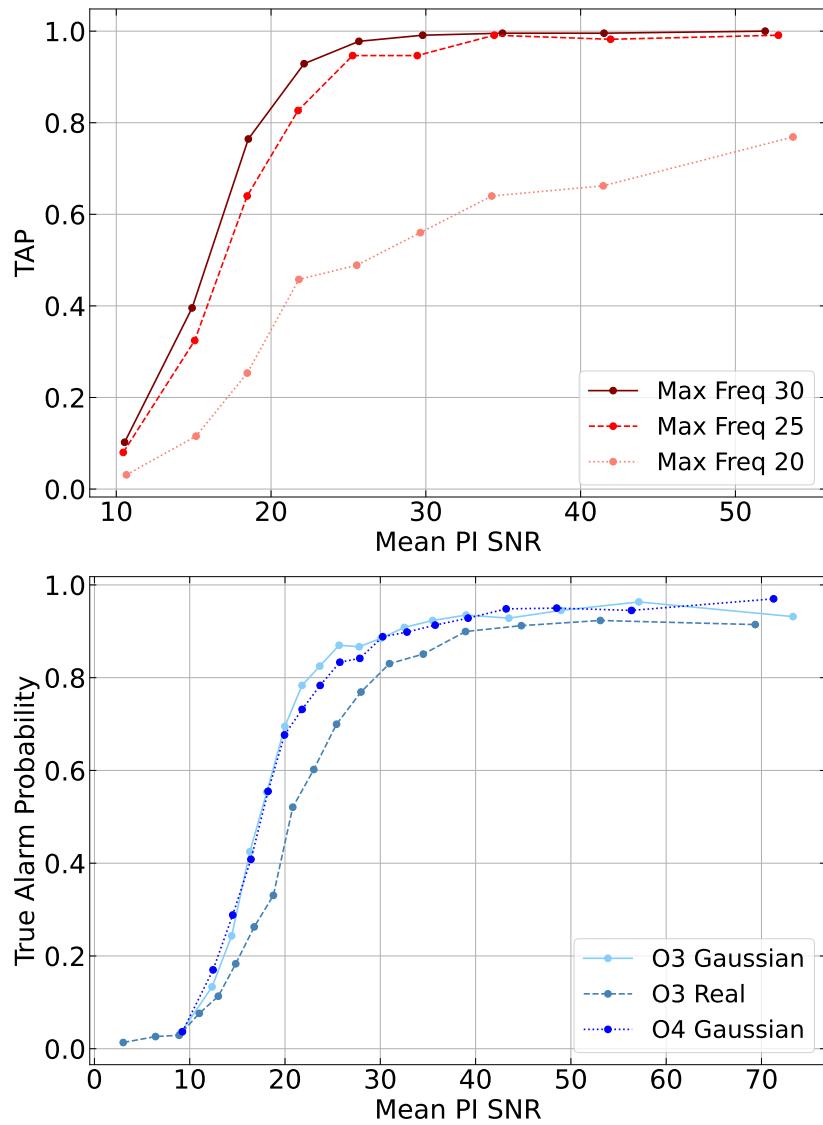


Figure 13.4: *Top:* The TAP as a function of the PI SNR for the O3 Gaussian noise case. Each curve represents a different test set with a different maximum frequency seen by the CNN. *Bottom:* The TAP as a function of the PI SNR for the O3 Gaussian noise, real O3 noise, and O4 Gaussian noise.

From one step to the other, we shift the window by 5 s. This is repeated until the full 1000 s are covered. Note that the step of 5 s is arbitrary and can be reduced for a realistic early-alert pipeline. The minimum step length should be the time required to load 300 s of data, pre-process it, and predict it with our network. The DL algorithm is fast and takes about 0.5 s on a CPU, and 0.01 s on a *GeForce GTX 750* GPU. The pre-processing is also fast: about 0.13 s to compute the PSD with PYCBC, 0.01 s to perform the whitening, and 1 s to remove the peaks and do the renormalization. The limiting factor is to load 300 s of data for three detectors with GWPY⁷, which takes around 2 s on an *Intel Xeon E5-2650 v4* CPU. Note that the PSD used for the whitening is computed each time we load the 300 s frame. To reduce the computational time further, one can evaluate the PSD at regular intervals and use the result for multiple steps.

Fig. 13.5 illustrates the time left before the merger when our approach detects the event for the different noise types. Each point contains 1000 frames with a duration of 1000 s, and each frame has a different noise realization. In each of them, we inject a BNS signal with fixed component masses similar to those detected for GW170817 [37]. This maintains a fixed duration for the data. The sky position of the signal varies from frame to frame. We then slide a 300 s window over the 1000 s of data, as described above. The process is repeated for injections corresponding to a larger distance. Fig. 13.5 shows that, for a given distance, the events are detected the earliest in O4 Gaussian noise. It is also interesting to note that the time before the merger for real O3 and Gaussian O3 noises is not that different, even if the Gaussian case is better. An event like GW170817 at a distance of 40 Mpc can be detected by our method 25 s in advance in real O3 noise, 35 s in advance in Gaussian O3 noise, and 50 s in advance for Gaussian O4 noise, showing quite good trigger capabilities for future observations runs.

Often, the performance of online matched-filtering searches is evaluated with a FAR. It represents the probability that a trigger occurs because of the noise for a given period of time [138]. The matched filtering FAR is computed for each event and represents how often the noise is expected to produce a trigger with a ranking statistic value at least as high as the one of the event. With our method, we can not compute such a FAR, but it is possible to evaluate the False Positive Over Time (FP_t), defined as the number of false positives for a given period.

To compute the FP_t , we run our network over the entire O3b data using the same setup as the one described previously. We shift the observation windows by 5 s for each step and veto the times corresponding to events reported in the GWTC-3 catalog [40]. Additionally, we assume that there are no other detectable events in the data⁸. Then, the FP_t is the number of triggers divided by the total observation time. For O3b, we obtain a FP_t of 277.54 per day, which is too high to be used for online searches. To decrease its values, we can consider that an event is present when our network gives multiple

⁷Using the built-in function `gwpy.timeseries.TimeSeries.get()`.

⁸This assumption is reasonable since our network needs relatively high SNRs to detect the inspiral, and the event would therefore have been detected by the usual search methods when merging.

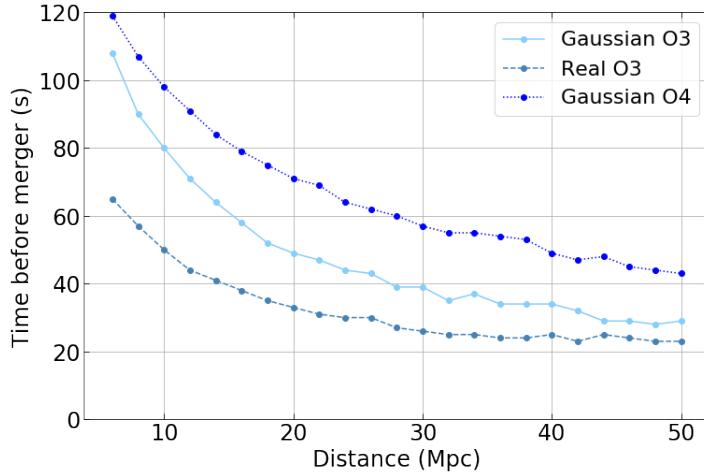


Figure 13.5: The time before the merger for the event’s detection as a function of the distance. This experiment is done for a BNS with component masses similar to GW170817.

triggers in a row, as shown in Fig. 13.6. If we keep detections with five consecutive alerts, the FP_t goes down to 12.31 per day, and it goes to 1.71 per day if we consider ten triggers in a row. Using multiple triggers implies longer delays before producing an alert and reduces the time before the merger for the detection. For example, considering five triggers leads to a delay of 20 s as we wait for 5 s for each step when sliding the window. In the end, we need to find a trade-off between the time before the merger and the desired FP_t .

Another way to decrease the FP_t is to use coherent triggers between two or more detectors. The training strategy for the network does not make it favor coherent triggers. Indeed, since it is trained for one, two, or three interferometers available, it learns to trigger even if only one detector is online. Furthermore, even if more than one detector is online, we do not use a minimum SNR in each of them for the training set. Hence, the network learns to trigger even if only one interferometer picks up the signal. In the end, it means that as soon as the CNN sees something remotely close to an inspiral in one of the detectors, it triggers, leading to a relatively high FP_t .

13.2.2 Estimation of the Number of BNS Inspirals Detectable in O4

To estimate the number of BNSs our network could detect in O4, we simulate a population of BNSs. It is generated using the method described in [319], and the BNS merger rate is normalized so that the local rate is equal to the median rate given in [326]. The only difference with [319] is that we adapt the detection thresholds and the PSDs to our O4 scenario. We keep BNS events with a network SNR higher than 13 and discard all the

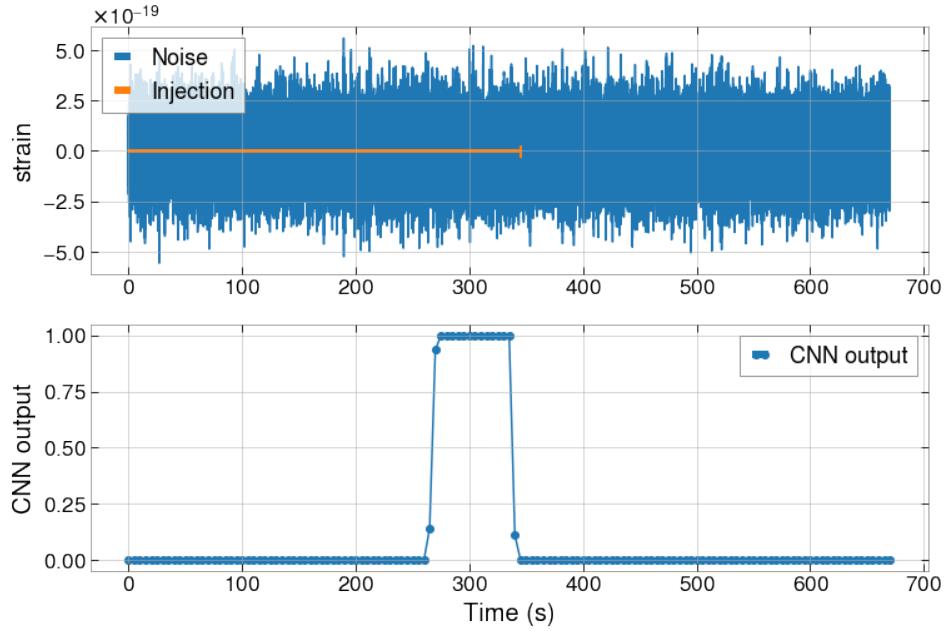


Figure 13.6: *Top:* Representation of a signal and the noise it is injected in. *Bottom:* Representation of the output of the CNN. Each point represents the probability of having an inspiral in the 300 s of data. By convention, the time of a point represents the end of the time window. The network does not trigger on the early inspiral because the PI SNR is too low. When it becomes high enough, the network produces a trigger until the injection leaves the frame, giving multiple points with a high probability in a row.

others. This threshold is chosen as we expect our network to find only BNSs visible in the detector network. A global SNR of 13 corresponds approximately to an SNR of 8 in each detector.

To have more statistics, we compute the equivalent of 5 years of data, with a duty cycle of 100% for all the detectors. Our simulations predict that, on average, around twenty BNSs per year will have a network SNR over 13 for O4 sensitivity. Our network can detect around three of those BNSs in advance. Fig. 13.7 represents the time before the merger for all the BNSs detected by our network over five years of data. Even if our network detects only three events out of twenty, it is important to note that these events are seen in advance and would, therefore, not be seen at that stage by the unmodified matched filtering searches. Nevertheless, matched filtering pipelines adapted to the early detection of long inspirals are also being developed [287, 334]. Those are also able to detect BNS mergers in advance. Even if the comparison between these works and ours is difficult (partially because of the difference in noise, but also performance evaluation), their times before the merger are comparable to those obtained by our network, ranging from $\mathcal{O}(10)$ to $\mathcal{O}(100)$ s. An advantage of these early-warning matched-filtering searches is that their FAR is lower than our FP_t (around one per month). However, they require

Time before merger	$\mathcal{M}_c (M_\odot)$	net SNR	net PI SNR at detection	Max. frequency
88 s	1.19	71.87	15.32	25.45 Hz
59 s	1.08	63.77	23.43	31.35 Hz
58 s	1.26	53.01	16.63	28.75 Hz
25 s	1.16	28.72	16.31	41.39 Hz
22 s	1.95	64.07	19.8	31.45 Hz
22 s	2.06	54.88	18.01	30.42 Hz
19 s	2.15	30.55	10.37	31.29 Hz
14 s	1.69	31.26	14.42	40.75 Hz
11 s	1.98	27.68	13.63	40.43 Hz
10 s	2.0	28.95	16.28	41.58 Hz
10 s	1.79	25.04	12.9	44.52 Hz
7 s	2.01	20.47	11.87	47.48 Hz
7 s	1.72	34.21	25.71	52.2 Hz
3 s	2.12	28.28	20.68	63.01 Hz

Table 13.2: The time before the merger, the maximum frequency seen by the network at detection, the chirp mass of the event, the network SNR, and the PI SNR at the moment of the detection for all the detected BNSs in five years of simulated O4 data.

more computational resources during the search as the highest cost for machine learning is moved to the training step. During the search, our method can run on a single GPU or even on a single CPU, while most standard matched-filtering methods require parallelization on multiple CPUs.

Table 13.2 shows the different characteristics of the detected BNSs. The network can see an event when the network PI SNR is between 10 and 25, which is expected according to Fig. 13.4. The time before the merger at which the CNN can detect a signal depends on two factors: i) the network PI SNR, and ii) the length of the signal. The PI SNR can be seen as a fraction of the SNR and its exact value depends on the signal part considered (hence the maximum frequency seen by the CNN). For a fixed signal duration, if the network SNR is high, the NN can detect an event at a lower maximum frequency, corresponding to a longer duration before the merger. However, if we fix the SNR and the maximum frequency while increasing the signal’s duration (for example by decreasing both the chirp mass and the luminosity distance), the event is detected earlier. This behavior is well represented in Fig. 13.7, where events with a low chirp mass and a high SNR are detected the earliest. It also explains why some events with a lower chirp mass can be detected earlier than events with a higher network SNR.

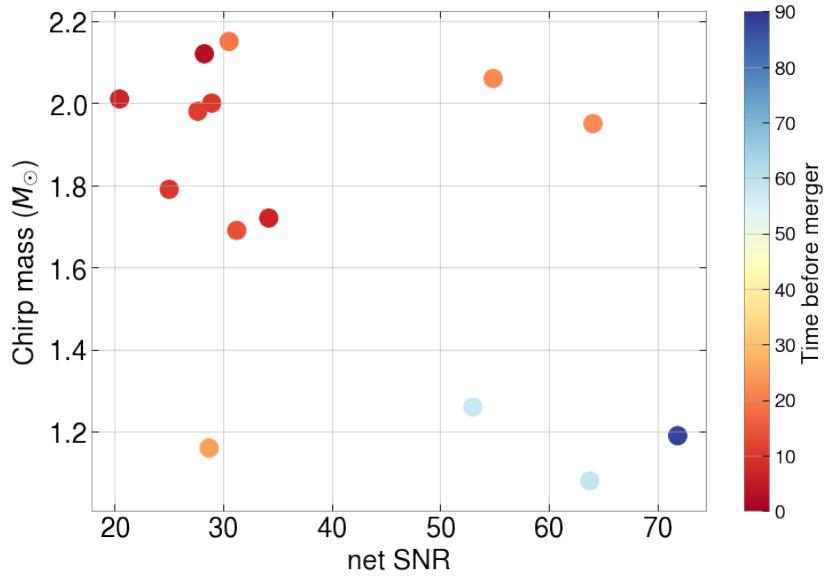


Figure 13.7: Chirp mass and time before merger as a function of the network SNR for the BNSs detected by our NN in five years of simulated O4 data. We can detect about three events in advance per year, up to a minute and a half before the merger.

13.3 Conclusions

This work builds upon the framework developed in Chapter 12. We implement several upgrades and modifications to the CNN-based pipeline designed to detect the early inspiral phase of BNS events. An important upgrade is the increased duration of the frames passed to the network, allowing us to search for smaller frequencies and opening the door to earlier detections. Another benefit of this increased duration is that we can use a single network to look for all BNS types, which was not the case in our previous work. The detection of events with smaller maximum frequencies is not easy and requires an adapted training methodology: curriculum learning. We consider realistic observation scenarios, including all the detectors of the LIGO-Virgo network, and use realistic noise realizations: O3 and O4 colored Gaussian noises, and real O3 noise. We have also demonstrated that, even in real detector noise, our network can identify GW signals in advance. We expect our network to see some BNSs up to minutes in advance in O4, provided the SNR of the event is high enough. In future work, we will upgrade our method to also search for NSBH mergers. As discussed in Sec. 13.2, we will also develop ways to decrease the FP_t . Finally, we will investigate avenues to infer the sky position with only the early inspiral part to also have an idea of where to look for an EM counterpart.

14

SWIFT SKY LOCALIZATION OF GRAVITATIONAL WAVES USING DEEP LEARNING SEDED IMPORTANCE SAMPLING

In addition to knowing when a BNS merger signal is incoming, it is also important to access its sky location. Indeed, one would need to localize it rapidly to extract a maximum of information. The best possible sky localization occurs when at least three detectors are online and seeing the event. Currently, low-latency sky location is done with BAYESTAR [292], a Bayesian algorithm leveraging the matched filtering information to by-pass the need for MCMC or nested sampling, and able to produce sky maps in less than a minute after detection. However, its sky location is not always robust compared to those obtained through complete Bayesian algorithms relying on MCMC or nested sampling.

Better early sky localization capabilities would improve real-time multi-messenger astronomy (MMA), observing astrophysical events through multiple channels - EM transients, cosmic rays, neutrinos - only seconds after the GW is detected. MMA is limited to GWs originating from BNS and NSBH mergers. According to current understanding, BBHs are unlikely to emit an EM counterpart during their merger [335, 336]. Currently, astrophysicists try to collect the non-GW channels in the weeks after the event. A notable example is GW170817 [37, 276]. This process takes enormous effort, while the obtained data quality is often sub-optimal. Having all channels observed for the entire duration of the event would be a major leap forward. Real-time MMA would enable a plethora of new science [278, 279, 337].

As mentioned before, real-time MMA relies on the generation of a skymap and imposes two limits on the methodology used to obtain it. First, it needs to be swift to allow observatories to turn towards an event's origin, preferably only seconds after its observation. Second, the skymap needs to be as accurate as possible since telescopes have limited sky coverage.

The most precise methods are the Bayesian inference methods such as LALINFER-ENCE [148] and BILBY [149]. They reconstruct the posterior distributions for all the compact binary coalescences (CBC) parameters. Given enough samples, they asymptotically tend to the true distributions [338]. However, such methods are slow, taking hours to days for BBH signals and up to weeks for BNSs. This makes them unusable for

real-time MMA.

Owing to its speed after training, ML represents an attractive alternative to traditional methods for producing rapid skymaps. Several methods have been developed to generate posterior distributions for CBC events using ML [293, 295]. Even if they rely on different approaches, these methods can generate a complete set of posteriors in about a few minutes. Both DL methods are fast and seem to be accurate for the 100 - 1000 simulated GW events they are evaluated on. However, these methods have a few issues: (i) they are both susceptible to changes in the PSD and SNR, (ii) both have performances close to Bayesian inference but do not match it, leading to different posterior distributions, (iii) they can act unpredictably outside of the trained strain-parameter pairs and, even within this space, they can act unpredictably due to the black box nature of NNs. Issues (i) and (ii) have been addressed for normalizing flow algorithms in [297]. However, the robustness remains behind those of traditional Bayesian inference.

In Ref. [300] and this chapter, we present a method trying to bridge the gap between Bayesian inference and DL methods, allowing for fast inference while preserving optimal accuracy. It is to be noted that combining Bayesian inference and DL methods has recently gained traction in the GW community, see [299] for example. The goal of our algorithm is to restrict the parameter space such that, via sampling, one can quickly obtain an accurate sky map. We use a multi-headed CNN to parameterize an independent sky and mass distribution for a given BBH event. The model is trained on simulated precessing quasi-circular BBH signals resembling the ones observed by the LIGO and Virgo detectors. The parameterized sky and mass distributions are Gaussian-like and are assumed to approximate the sky and mass distributions generated by Bayesian inference. Using the parameterized sky and mass distributions, we construct a proposal posterior in which all other BBH parameters are distributed uniformly. By using importance sampling we can then sample from the exact reference posterior. This implies that we effectively match the performance of Bayesian inference in a short time without exploring the entire parameter space. We stress that this work is a proof-of-concept to show the promises of combining NNs and Bayesian inference. In future work, we will consider more complex DL structures and other event types.

14.1 Methodology

Our inference setup is a two-step method. In the initial step, we infer simple distributions for the BBH sky localization and masses by using a NN. Subsequently, we apply importance sampling (see Sec. 6.4.3) to these simple distributions to compute a more accurate posterior.

14.1.1 Going from Simple Distributions to Complex Posteriors

High-dimensional distributions in which the majority of the probability density is confined to a small volume of the entire space are hard to sample from, which results in long run times to get proper estimates when using MCMC methods. A well-known avenue to cope with this problem is importance sampling (see Sec. 6.4.3). In this case, it suffices to have a good proposal distribution and a likelihood to generate the final posterior distributions.

Generating accurate posteriors for GW observations using MCMC is very time-consuming, and thus importance sampling is an attractive alternative. Importance sampling requires us to have a viable proposal distribution. Published posteriors for known GWs show that the probability density in the posterior is relatively confined for both the sky location and the two masses [267]. A Von Mises Fisher (VMF) and Multi-Variate Gaussian (MVG) distribution are good first-order approximations of the sky and mass distribution, respectively, and thus suitable to use as a proposal distribution for importance sampling. We propose to construct this proposal distribution by assuming a uniform distribution over all non-spinning BBH parameters, except for the sky angles, represented by a VMF, and the masses are represented by an MVG distribution. Assuming that the sky angles, masses, and other BBH parameters are independent, our proposal distribution becomes the product of these two distributions. In what follows, we discuss how we create this proposal distribution using a NN.

Importance sampling also requires a likelihood function for the proposal distribution and the desired distribution. In the previous paragraph, we discussed how we want to create a proposal distribution. The desired distribution is the likelihood function for GWs given in Eq. (4.18).

We now have all the parts needed to discuss how we utilize importance sampling for a given strain s . A trained neural network parametrizes the proposal distribution q for the given strain. The proposal distribution generates n samples representing possible GW parameter configurations. For each sample, we calculate the logarithm of the importance weight $w(\boldsymbol{\theta})$,

$$\log w(\boldsymbol{\theta}) = \log \Xi(s|\boldsymbol{\theta}) - \log q(\boldsymbol{\theta}) + C, \quad (14.1)$$

where $\boldsymbol{\theta}$ is the BBH parameters and Ξ is the right-hand-side in Eq. (4.18), instead of the importance weight itself to prevent numeric under- and overflow. The constant C is added to set the highest $\log w(\boldsymbol{\theta})$ to zero to prevent numerical errors. Since we normalize the weights afterward, the correct importance weights are obtained in the end. The reweighed samples represent the desired distribution p .

If the proposal distribution does not cover the true distribution well enough, the importance samples are dominated by a single to a few weights if we restrict the run-time. We can use this as a gauge to check if the skymap produced by the NN and importance sampling can be trusted.

14.1.2 Neural Model

Previous work [312] shows that CNNs can extract the masses from a BBH event just as well as the currently-in-use matched filtering. Furthermore, another work [339] indicates that 1D CNNs can locate GW origins. Therefore, we chose to use a 1D CNN to model the distribution across the sky for the source and a multivariate normal distribution for the two masses of the BBH system.

The network architecture of this 1D CNN is presented in Fig. 14.1 and consists of four parts: a convolutional feature extractor and three neural network heads. These heads are used to specify the two distributions. The following properties were tested or tuned for optimal performance: the number of convolutional layers, kernel size, dilation, batch normalization, and dropout. The model shown in Fig. 14.1 produced the best result on a validation set.

The convolutional feature extractor generates a set of features characterizing a given GW. This set of features is passed to the neural heads. Each head is specialized to model a specific GW parameter. The first head determines the sky distribution, the second head the masses, and the third head the uncertainty over the two masses. Below we elaborate on each of these heads and how they characterize these distributions.

The first head specifies the distribution of the source position. A 2D Gaussian distribution is ill-fit because the sky location is a surface on a 3D sphere. A suitable alternative is the VMF distribution [340] which is the equivalent of a Gaussian distribution on the surface of a sphere. The probability density function and the associated negative log-likelihood (NLL) of the VMF distribution:

$$p(x|\mu, \kappa) = \frac{\kappa}{4\pi \sinh(\kappa)} \exp(\kappa x^T \mu) \quad (14.2)$$

$$\text{NLL}_{\text{VMF}}(x, \mu, \kappa) = -\log(\kappa) - \log(1 - \exp(-2\kappa)) - \kappa - \log(2\pi) + \kappa x^T \mu, \quad (14.3)$$

where x and μ are normalized vectors in \mathbb{R}^3 , with the former being the true direction and the latter being the predicted direction. κ is the concentration parameter, which determines the width of the distribution. It plays the same role as the inverse of the variance for a Gaussian distribution. We use this distribution by letting the first head output a three-dimensional vector $D = (D_x, D_y, D_z)$. The norm of D specifies the concentration parameter κ , and its projection onto the unit sphere gives the mean μ . So, $\kappa = |D|$, and $\mu = D/|D|$. These values, together with the true direction x , are used to calculate the negative log-likelihood, which is used as a loss function for the first head.

The second and third neural heads specify a 2D MVG, describing the possible configurations of the masses. The means ν of the MVG are given by the second head, and the covariance matrix Σ is specified by the third head. Given the true values of the masses $y = (m_1, m_2)$, the probability density function and associated negative log-likelihood of

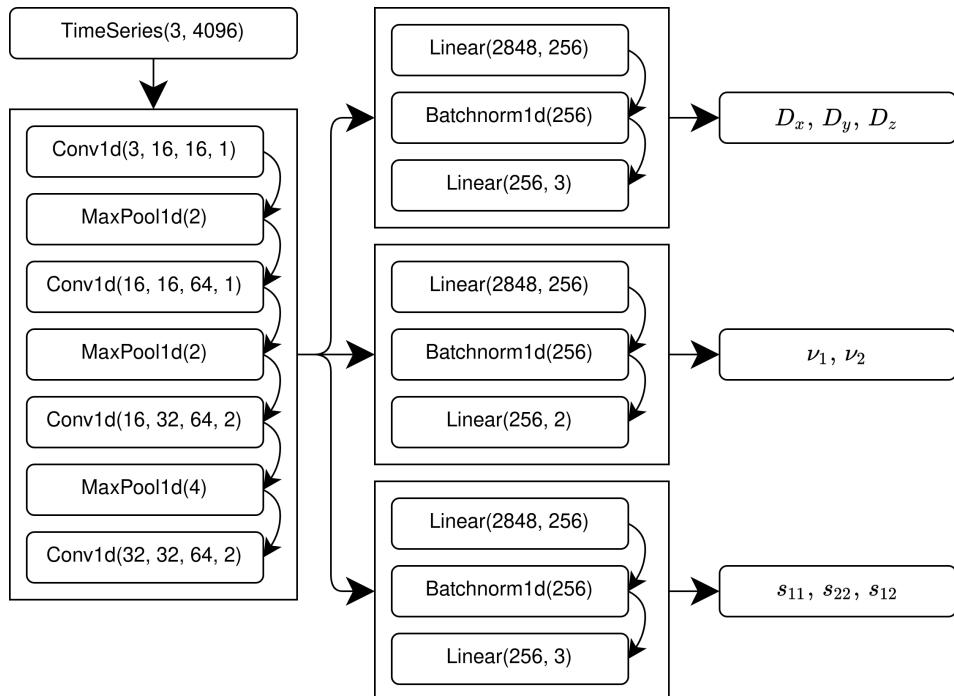


Figure 14.1: A graphical depiction of the CNN developed in this work. After each MaxPool1d and Batchnorm1d layer, a leaky ReLU activation function with an $\alpha = 0.1$ is applied. The convolutional part is shown on the left and takes as input a time series of 4096 elements with 3 channels. $\text{Conv1D}(i, o, k, d)$ denotes a 1D convolution with i input channels, o output channels, kernel size k and dilation factor d . $\text{MaxPool1d}(k)$ denotes a 1D max pooling layer with kernel size k . The output of the convolutions is given to three independent NN heads. The first head predicts the sky location parameterized as $D = (D_x, D_y, D_z)$, the second head predicts the mean of the masses of the two black holes, and the last head predicts the uncertainty elements of the covariance matrix over the two masses. $\text{Linear}(i, o)$ denotes a linear transformation with i input features and o output features. Lastly, $\text{Batchnorm1d}(i)$ denotes a 1D batch normalization layer with i input features.

the MVG are:

$$p(y|\nu, \Sigma) = \frac{1}{\sqrt{(2\pi)^2 |\Sigma|}} \exp\left(-\frac{1}{2}(y - \nu)^T \Sigma (y - \nu)\right) \quad (14.4)$$

$$\text{NLL}_{\text{MVG}}(y, \nu, \Sigma) = \frac{1}{2}(y - \nu)^T \Sigma^{-1} (y - \nu) + \frac{1}{2} \log(|\Sigma|) + \log(2\pi) . \quad (14.5)$$

The inverse covariance term in the negative log-likelihood can contain imaginary numbers if the covariance matrix is not positive-definite. To ensure the covariance matrix Σ remains positive-definite, it is parameterized through

$$\Sigma_{11} = \exp(s_{11}) \quad (14.6)$$

$$\Sigma_{22} = \exp(s_{22}) \quad (14.7)$$

$$\Sigma_{21} = \Sigma_{12} = \tanh(s_{12}) \sqrt{\Sigma_{11} \Sigma_{22}} . \quad (14.8)$$

The three variables s_{11} , s_{22} , s_{12} are predicted by the third neural head and define the covariance matrix completing the MVG prediction of the masses. The parameterization and implementation of the MVG are based on Ref. [341].

By further assuming that the sky and mass distributions are independent, we obtain a first approximation of the posterior distributions, thereby satisfying the requirements for importance sampling.

14.2 Setup for the Experiments

Experiments were performed on two different fronts: (i) training the NN followed by the empirical evaluation of its performances on unseen test data, and (ii) comparing the NN model, importance sampling scheme, and BILBY, based on several metrics and sky maps. Below, we describe the experimental details and justify the decisions we made. All experiments are performed on a computer with a 16-core *AMD Ryzen 5950X* CPU, *NVIDIA 3090 RTX* GPU, and 64 GB of RAM. The source code is available at <https://github.com/akolmus/swiftsky>.

14.2.1 Setup for Training and Evaluating the Neural Network

To obtain strain-parameter pairs for training and validation, we sample parameters from a BBH parameter prior (see Table 14.1) and generate the associated waveforms using the *IMRPhenomPv2* waveform model [107]. The waveforms are generated in the frequency domain in a frequency band of 20 to 2048 Hz. The duration of the signal

parameter	prior	minimum	maximum	unit
Masses (constraint)	-	20	80	M_{\odot}
Chirp mass	Uniform	10	100	M_{\odot}
Mass ratio	Uniform	0.25	1.0	-
Spin magnitudes	Uniform	0	0.95	-
Spin polar angles	Sine	0	π	rad
Spin azimuthal angles	Uniform	0	2π	rad
Right ascension	Uniform	0	2π	rad
Declination	Cosine	-0.5π	0.5π	rad
Binary inclination angle	Sine	0	π	rad
Coalescence phase angle	Uniform	0	2π	rad
Polarization angle	Uniform	0	2π	rad
Time Shift	Uniform	-0.1	0.1	s
Luminosity distance	-	1000	1000	Mpc

Table 14.1: The priors used for the data generation when training and testing our NN. The luminosity distance in the prior was set to 1000 Mpc and scaled afterward to match the desired SNR.

is 2 s. Subsequently, these waveforms are projected onto the LIGO-Hanford, LIGO-Livingston, and Virgo interferometers. We sampled the SNR from a scaled and shifted beta distribution with its peak set to 15 (see Figure 14.2). The luminosity distance in the prior is set to 1000 Mpc and scaled afterward to match the desired SNR. We generate Gaussian noise from the design sensitivity PSD for each detector [25, 26]. Finally, the signal is injected into the noise, and an inverse Fourier transform is applied to obtain the strains as time series. This setup allows us to generate an arbitrary amount of unique strain-parameter pairs, which results in every training epoch having a unique data set, reducing the risk of over-fitting.

We applied three preprocessing steps to the data. All time series are whitened with the aforementioned PSDs. Next, the time series are normalized. Lastly, to make the mass distribution easier to learn, we calculate a shift and scaling factor for the target masses such that all target masses are between -1 and +1. The shifting and scaling are applied inversely to the NN output during importance sampling to get the correct masses.

The model is trained for 300 epochs with a batch size of 128. During each epoch, we drew 500 000 strain-parameter pairs for training and 100 000 strain-parameter pairs for validation. Adam [229] is used to optimize the model in conjunction with a cosine annealing scheme with warm restarts [342]. The learning rate oscillates between 10^{-3} and 10^{-5} with a period of 20 epochs; weight decay is set to 10^{-6} . Multiple hyper-parameter configurations are tested, and this configuration obtained the best performance.

To benchmark the trained model, an unseen test set is generated. It contains 100 000 strain-parameter pairs at specific SNR values. The model is evaluated using the mean

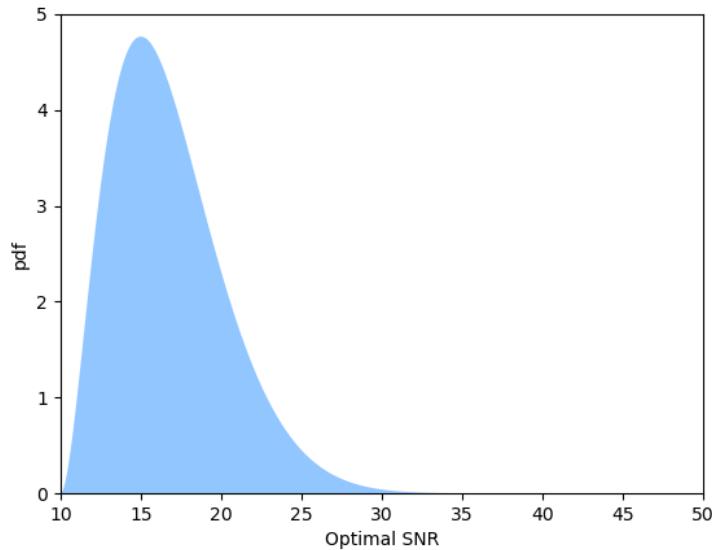


Figure 14.2: Scaled and shifted beta distribution acting as the SNR sampling distribution during training and validation. The vertical axis represents the probability density function of this beta distribution. The horizontal axis represents the SNR value.

absolute angular error (maae) and the average 90% confidence area of the predicted VMF distribution.

14.2.2 Setup for Importance Sampling

To evaluate the importance sampling procedure, we construct a slightly simpler test set in which we set the spin magnitude to zero. This is done to limit the BILBY run-time. The importance sampling procedure is applied to the first 100 strain-parameter pairs of this test set at three different optimal SNR values: 10, 15, and 20. For each strain-parameter pair, we generate 200 000 importance samples. To simulate multiple independent runs at various time points for the same strain-parameter pairs, we sub-sample from these 200 000 importance samples during the experiments.

We ran two experiments to test the convergence of the importance sampling method. In the first one, we use the importance sampling scheme as a maximum likelihood estimator. For a given set of samples, we choose the sample with the highest likelihood and calculate the angle between this sample and the true sky coordinates. In the second experiment, we represent the probability density function of the importance samples by a KDE and test how well the resulting density covers the injected right ascension.

Specifically, we use a Gaussian KDE¹ to fit the right ascension distribution proposed by the importance samples. The log-likelihood of the actual right ascension is used to measure the quality of the estimated density. We remove a few outliers from the second experiment. By restricting ourselves to only the right ascension, the number of outliers decreases. These outliers have densities that do not cover the true right ascension at all, resulting in extreme negative log-likelihoods dominating the average log-likelihood. For both experiments, we expect the metric to improve as the number of importance samples increases, and to level after a significant number of importance samples indicating convergence.

14.2.3 Generating Skymaps

We use BILBY as a benchmark to generate skymaps for the first ten strain-parameter pairs of the test set and, for each, create a version at an SNR of 10, 15, and 20. To make a fair comparison, the prior given to BILBY has its spin components set to zero. Moreover, the posterior inference is done with standard settings, and each run took between 2.5 and 7 hours to complete. During these runs, the live points of the sampler are saved every 5 seconds and labeled by the total number of sampled points. These points are used to run the two importance sampling experiments for BILBY.

14.3 Results

In Fig. 14.3, we summarize the results for the first experiment: the top panel gives the maae in the sky location, and in the bottom one we plot the 90% confidence area of the VMF distribution. As expected, when the SNR increases, the prediction error in the sky location decreases and the 90% confidence area becomes smaller. The error in the mass prediction is similar to those of other CNN approaches [312], see Fig. 14.4, indicating that the setup works well. We note that the error in the sky location seems quite high for $\text{SNR} < 10$ and does not converge to zero for high SNR. We can think of two possible explanations for the poor performance at low SNR. First, the detection rate using either CNNs or matched filtering pipelines at an SNR of 5 is less than 40% [312, 343]. At such a low SNR, it is difficult for the model to discern the differences in arrival time at each detector, which explains the slightly better than random predictions for $\text{SNR} < 7$. When we compare our angular error with other CNN approaches [339, 344], the average error looks similar. Furthermore, Ref. [345] reported that Gaussian approximations are only accurate for high SNR ($\text{SNR} > 8$), and, even then, multi-modalities might arise. Second, the sky distribution can be multimodal. This multi-modality is either due to

¹We use the *gaussian_kde* from the SCIPY python package.

strong noise or can be due to a sky reflection [148]. For three detectors, there are two viable solutions to the triangulation problem: the true sky location and its reflection. In most cases, the amplitude information is sufficient to break the degeneracy between the location and its reflection. However, for certain sky angles, this amplitude information does not lift the degeneracy, and a multimodal distribution is observed. For these angles, the model has a 50% chance of guessing the wrong mode and thus has an average angular error of 90°.

14.3.1 Importance Sampling

The results of the importance sampling experiments are shown in Fig. 14.5. The top panel shows the maae as a function of the number of importance samples. We observe a decreasing error when the number of samples increases. The bottom panel shows the log-likelihood of the true right ascension given by KDEs built using varying numbers of importance samples. The majority of the convergence in the maae seems to happen within the first 30 000 samples. The slow convergence can largely be attributed to strains for which the model predicted a wide sky distribution. When we compare this to results of BILBY, we see that the maae of the highest likelihood sample for all SNR is always between 1° and 8°. Importance sampling is competitive for an SNR of 20 and close for an SNR of 15, especially when considering that, in both cases, 2 out of the 100 sky distributions were parameterized as the sky reflection.

However, importance sampling is not competitive with BILBY in the second experiment. For all SNR values, BILBY reports log-likelihoods between 2 and 3, see the top panel of Fig. 14.6, and importance sampling does not reach these values. If we consider runs showing good convergence, *i.e.* where 90% of the importance weight is not determined by less than ten importance samples, importance sampling also reports log-likelihoods between 2 and 3. In the bottom panel of Fig. 14.6, we have repeated the kernel density experiment, but only for the well-converged runs. These runs represent 30% of all runs, and almost no $\text{SNR} < 10$ runs.

14.3.2 Generating Skymaps

As a final test, we generate sky maps using the NN, importance sampling, and BILBY on the same signals. Three representative sky maps are shown in Fig. 14.7. The sky maps generated by the NN have significantly more spread than those generated by importance sampling and BILBY. As we explained in the previous sections, this might be due to the NN overestimating the uncertainty and having difficulty extracting the exact signal from the detector noise. The sky maps generated by importance sampling and BILBY resemble each other quite a lot, their peak intensities are in the same position, and the

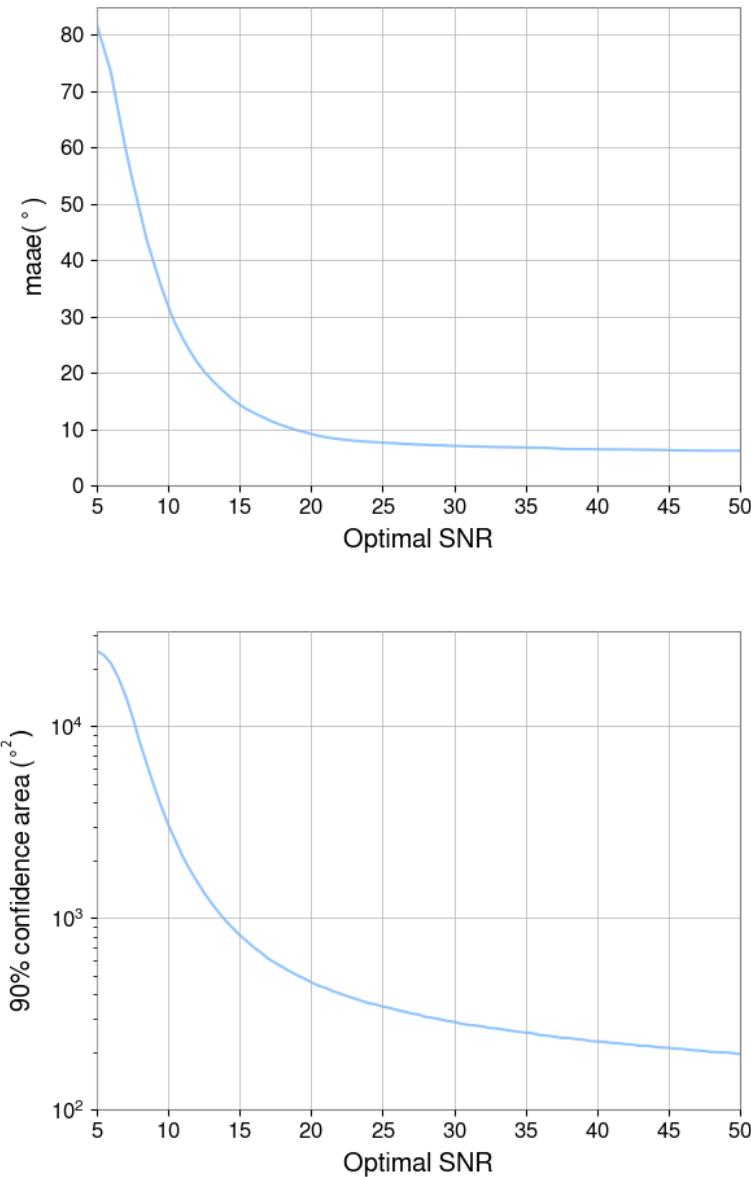


Figure 14.3: Characterization of the neural network in terms of accuracy and certainty over the test set. *Top:* the maae between the sky angle predicted by the model and the actual sky location as a function of the SNR. *Bottom:* the average size of the 90% confidence area, expressed in degrees squared, of the predicted VMF distributions as a function of the SNR.

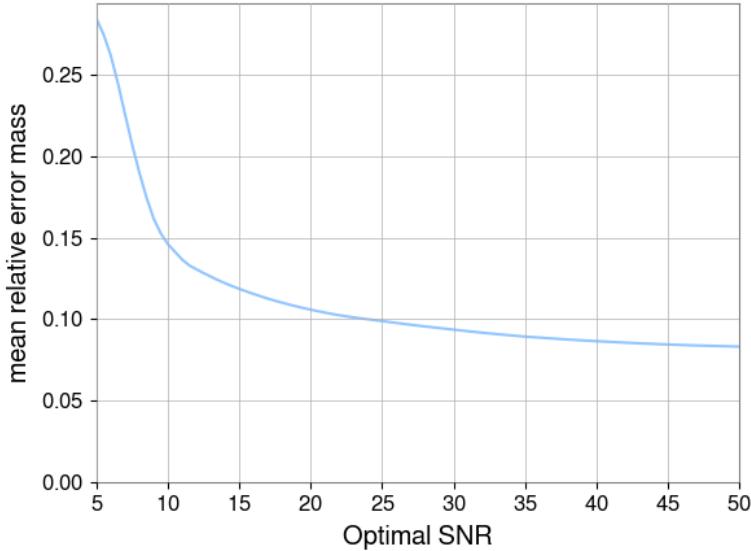


Figure 14.4: The mean relative error of the estimated component masses by the NN on the test set as a function of the optimal SNR. This figure closely resembles Fig. 5 in [312]. Any differences are due to the difference in the experimental setups. The main differences are that our priors include spins and that we do not use a stationary sky origin.

sky distributions occupy roughly the same area. However, the importance sampling sky maps are grainy and sometimes do not cover the complete area that BILBY does. As seen in the bottom row of Fig. 14.7, when the predicted VMF distribution has its peak intensity in the correct location the importance sampling creates better-looking sky maps. This improvement is due to the increased number of significant importance samples. These results indicate that a larger number of significant importance samples is needed, which is to be expected with only 5 min of run-time. Within only 1-4% of the BILBY run-time, we can already recover the essentials of the sky maps. Additionally, such sky maps can be good enough for low-latency searches and can be updated as more importance samples are generated.

14.4 Conclusions

In this study, we produced sky maps for simulated BBH events using an importance sampling scheme that turns an approximate sky map made by a NN into a sky map representing the Bayesian posterior distribution. Experiments show that our method is competitive with BILBY and can produce the essentials of the sky map within 4% of the BILBY run-time. However, in some cases, the proposal distributions made by the

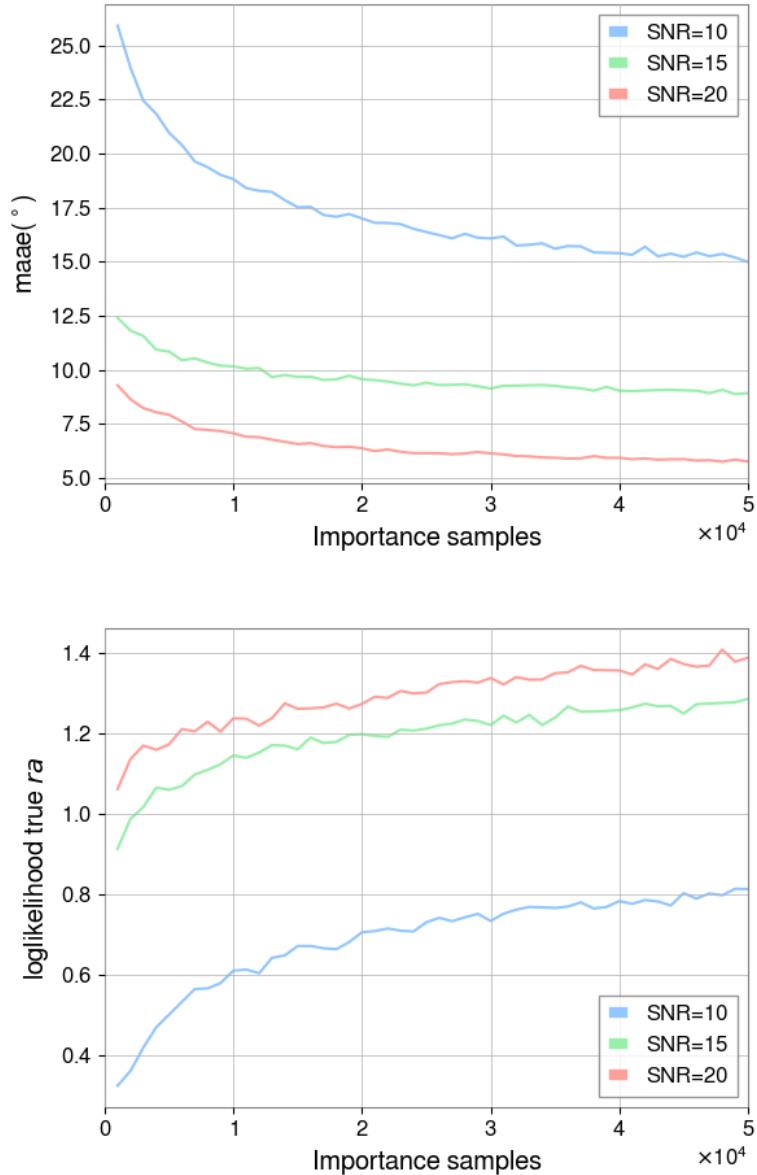


Figure 14.5: Characterization of the importance sampling, with the number of importance samples ranging from 1 000 to 50 000. The colors represent different SNR values, with blue, green, and red being 10, 15, and 20, respectively. *Top:* the maae of the importance sample with the highest likelihood as a function of the sample size. *Bottom:* the log-likelihood of the true right ascension according to the kernel density estimator created by importance samples as a function of sample size.

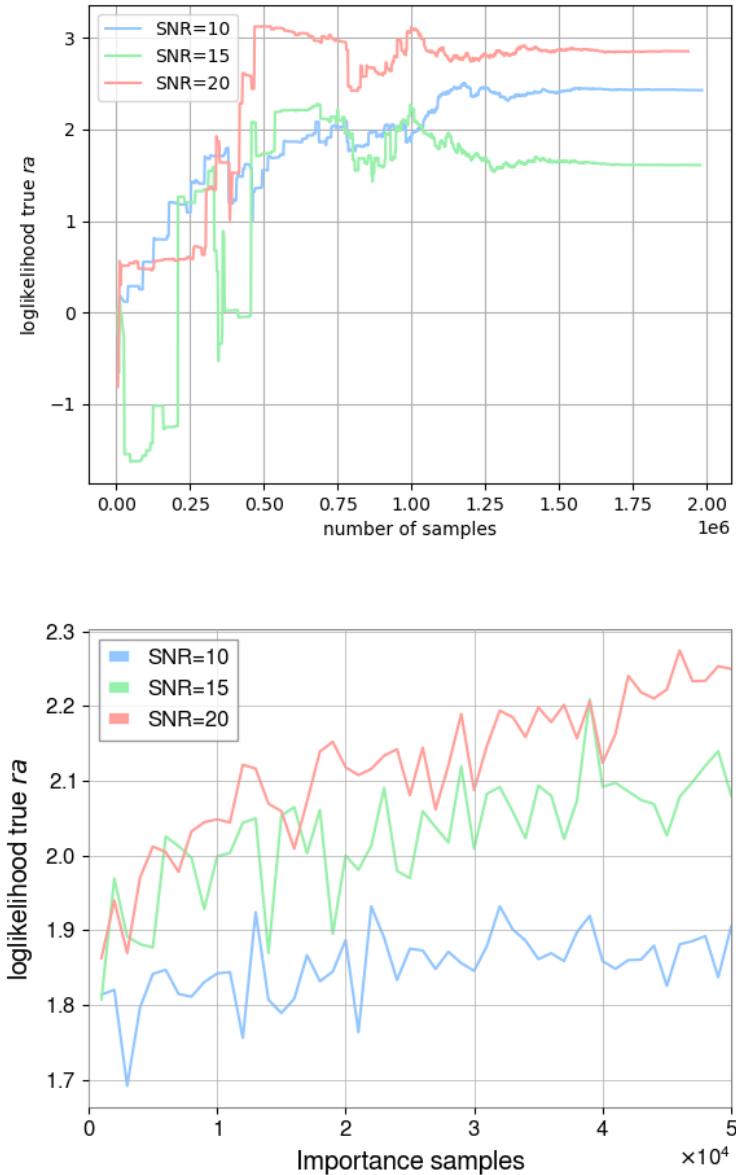


Figure 14.6: *Top:* The log-likelihood of the true right ascension according to the KDE created by the BILBY samples. The horizontal axis represents how many samples BILBY has generated (live and dead samples). *Bottom:* The log-likelihood of the true right ascension according to the KDE using only the importance samples of well-converged runs. These values are more in line with those of BILBY (see top panel).

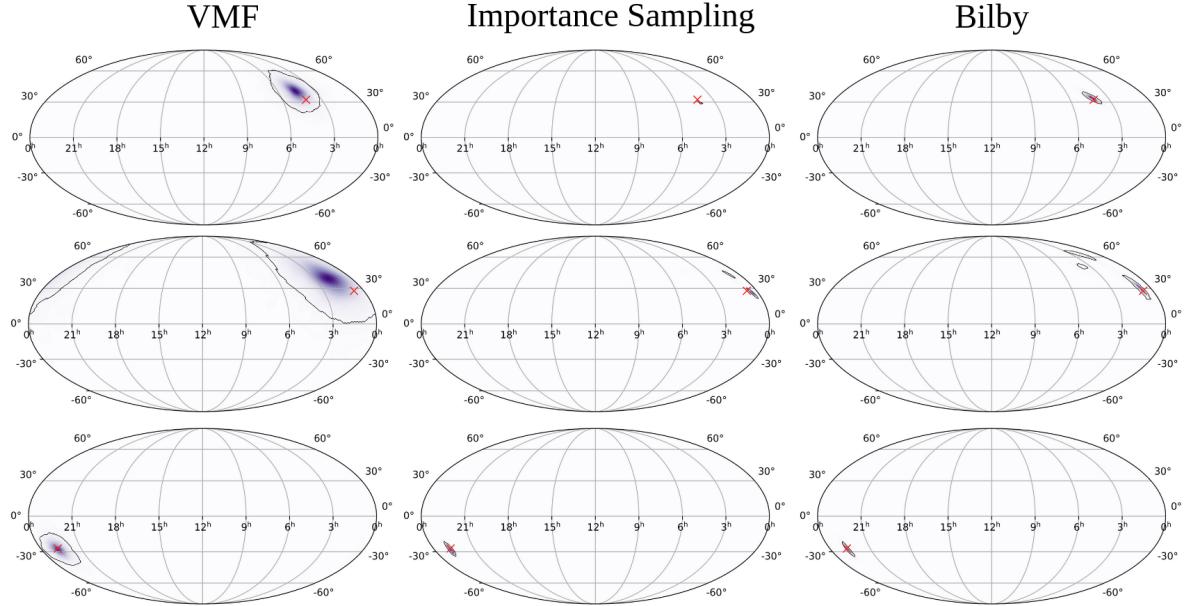


Figure 14.7: Examples of predicted sky maps by our NN (left), importance sampling after 100 000 steps or roughly 5 min of computing time (middle), BILBY at convergence (right). The true sky location is indicated in red. The shown sky maps are generated for signals with an SNR of 15. The number of significant importance samples, and hence the quality of the sky maps, increases as we go from the top to the bottom row. In the worst case scenario, the sky location is wider than for BILBY. However, it seems good enough to perform prompt follow-ups. In the best-case scenario, the sky location obtained after importance sampling is equivalent to the one given by BILBY.

NN are too crude, which hampers the efficiency of the importance sampling scheme. If the sampling efficiency is improved, importance sampling could be used as a quick alternative to BILBY or LALINFERENCE for inferring the GW posterior. In future work, we will consider more advanced DL models, such as normalizing flows, to infer more accurate posterior distributions. Additionally, we will accommodate other signal types and try to use our framework for early sky location generation.

This framework combined with the one presented in Chapters 12 and 13 would make for a complete low-latency/early-alert ML-based framework. Still, this would require some updates to the two pipelines. For example, the early-alert system should have a reduced FAP. This can be done by using consistency requirements between the triggers throughout the detectors. On the other hand, the sky location algorithm leads to prompt sky localization but does not produce a sky map in advance. Therefore, we would need to adapt the model to produce sky maps based only on the inspiral. However, we have also seen that our algorithm does not perform well on low SNR signals. Therefore, we would probably need to modify our NN to deal properly with the inspiral part of the signal. Once both pieces work individually for more realistic scenarios, we can combine them to make a complete early-alert system.

Part IV

THIRD-GENERATION DETECTORS:
INTERESTING SCIENCE,
TOUGH ANALYSES

15

THIRD-GENERATION DETECTORS: EXPENDING OUR HORIZONS

Even if they are upgraded continuously and should reach higher sensitivities in the coming years [25, 26], the 2G detectors have inherent limitations due to their design and size. Therefore, planning for future ground-based interferometers has already started, with Einstein Telescope (ET) on the European side [30, 31] and Cosmic Explorer (CE) on the American side [50]. These detectors are designed to have a significantly increased sensitivity through various technological and design improvements.

Since the detectors are not in the commissioning phase yet, discussions on their exact design are still ongoing. For ET, the proposed plan is to make a triangular detector composed of nested interferometers with a 60° opening. The triangle would have a 10 km arm-length with a “*xylophone*” structure, meaning that it would be made of two entangled detectors, one specialized for lower frequencies and one for high frequencies [31]. The triangular shape gives a more uniform antenna response, making for better sky coverage. Additionally, it leads to redundancy and makes the construction of a null-stream possible [346–348]. For the low-frequency part, the plan is to use cryogenic mirrors and low-power lasers. On the other hand, the high-frequency detector should have a high-power laser to reduce the shot noise [31, 31, 349]. Finally, to avoid gravity gradients and perturbations due to anthropological activity, the full structure would be placed underground. With these nested detectors, it would be possible to reach frequencies between 1 and 5 Hz [31]. The latter is crucial for the high-mass CBC mergers detection, as higher masses lead to lower maximum frequency, as seen in Eq. (2.63). On the other hand, the generally improved sensitivity makes the LIGO BBHs visible to redshifts of ~ 20 . At that time, no stars had formed. Therefore, only primordial BHs would be observed [52]. Moreover, BNS mergers should be visible up to redshifts of 3. This is after the peak in the star formation rate. Therefore, the number of detections should increase drastically compared to 2G detectors [52]. This is the theoretical design and studies are still ongoing to see the consequences of the setup.

In synergy with ET, CE should also be operational. Its design is different from ET’s. CE is expected to keep the L-shaped structure but is expected to have 40 km long arms [50]. It should also come with more powerful lasers to increase its sensitivity. However, nothing particular is planned to improve the low-frequency detection capacities. For 2G sources, it should reach the same type of horizon distances as ET, if not a bit higher thanks to its longer arms.

Together, these detectors could shed light on many open questions, see Ref. [52] for an

overview. It ranges from the origin of BBHs to the possible existence of primordial BHs. It also has interesting applications for BNSs, where many multi-messenger detections could be made [52, 285], and the high-frequency sensitivity makes the post-merger part of the signal observable [350–352]. The latter could help study the NS EoS as it happens in different conditions compared to the inspiral.

Besides the technological challenges, 3G detectors also require major developments on the data analysis side. For CBC mergers, the enlarged horizon leads to an enormous number of mergers happening over time, with up to hundreds of thousands of CBC signals observed over a year [52]. This means the analysis tools will have to be fast if one wants to keep up the pace of the analysis. Additionally, the improved low-frequency sensitivity significantly increases the duration of the signal, as can be seen from Eq. (12.4). While BBHs will last for minutes, BNSs will last for hours. This is interesting for early alert pipelines as BNS systems could be very loud – with SNRs of a few hundred –, and the SNR has time to accumulate in the first hours, and detectability could be reached well before the merger. On the other hand, the full signal analysis, from early inspiral to merger, will be computationally challenging. Combining this with the rate, current analysis methods [148, 149] will probably be unable to follow up on all the events. One can think of various techniques to circumvent these issues. Some examples are adaptive frequency resolution [353], relative binning [354–357], machine learning [293–299], quantum computing [358], … However, most of these methods are not yet mature enough to be used under realistic conditions.

Other problems will arise. Among them are overlapping signals [319, 359–363], where the combined increase in detection rate and signal duration leads to several signals in-band at the same time. They can interfere and could be problematic for data analysis in 3G detectors. This part of the thesis takes a look at this problem and offers several solutions.

16

BIASES IN PARAMETER ESTIMATION FROM OVERLAPPING GRAVITATIONAL-WAVE SIGNALS IN THE THIRD GENERATION ERA

When going from 2G to 3G detectors, there will be a steep increase in the detection rates. Moreover, the decrease in low-frequency sensitivity will lead to longer in-band signals [364]. These two combined effects lead to a significant probability for the signals to overlap, as first pointed out in [359]. In Ref. [319] and this chapter of the thesis, we study the occurrence of overlapping signals and their consequences on parameter estimation.

It is important to know how much overlapping signals can impact some of the 3G science goals (for examples, see [52, 364–373]). Not only the CBC science objectives can be affected but also other scenarios where one needs to subtract foreground sources, such as searches for primordial BHs [374–379]. In principle, individual CBC detections should not be impacted significantly by the occurrence of overlapping signals [380]. However, their recovered posterior distributions could be modified because of the overlap.

While previous work has focused on the study of individual sources in the 3G era [381–383], we here focus on the possibility to observe bias in the recovered parameters when neglecting the presence of another GW signal in band. First, we study how often signals overlap over a year in a 3G scenario with a network made of two CEs and one triangular ET detector. Therefore, we assume some realistic merger rate distribution as a function of redshift and mass distributions for BBHs and BNSs. We find that tens of CBC merger signals occur during one BNS signal. Also, events merging within the same second will be quite common. Since these are the cases where the largest bias is expected, we focus on them to simulate two closely-merging signals injected in simulated noise from the ET-CE network. We then analyze the signals using state-of-the-art parameter estimation techniques. We explore different types of overlaps: two BBH signals, two BNS signals, and a BBH with a BNS. In our simulations, we choose signal parameters consistent with past observations for the different types of sources. BNSs are modeled with parameter values similar to the ones observed for GW170817 [37], high-mass BBHs are represented by GW150914-like parameters [33], and lower-mass BBHs with GW151226-like parameters [384]. This work finds that, in most cases, the intrinsic parameters are recovered with negligible bias. However, when the merger times

of the signals are sufficiently close, considerable bias can occur when a short BBH or a quieter BNS overlaps with a louder BNS signal. Even if this work is exploratory, it already shows where issues can occur and where current parameter estimation techniques should be improved.

16.1 Overlap Rate Estimates

Before looking at the impact of overlapping signals on parameter estimation for the individual signals, we want to address the question of how frequently such overlaps occur, depending on the signal’s type.

16.1.1 Methodology

Previous characterizations of the overlap probabilities for 3G detectors were based on the *duty cycle*, defined as the ratio of the typical duration of a particular type of event (BNS or BBH) to the average time interval between two successive events of that type, assuming some fixed canonical values of the component masses for each type [359]. However, here we also want to allow for overlaps of mixed types and a range of component masses (and hence signal durations) within a given type to arrive at a detailed assessment of overlap rates. Therefore, we assume particular merger rates as a function of redshift for BBH and BNS, as well as component mass distributions, and based on these create simulated “catalogs” of signals in the detectors. This allows us to make quantitative statements regarding BNS signals overlapping with other BNS signals and with BBHs, and the same for overlaps of BBH with BBH events, in a much more detailed and realistic fashion¹.

We start by estimating the number of individual BBH and BNS coalescences happening in a given volume up to a maximum redshift, chosen to be $z_{max} = 30$ for BBH events and $z_{max} = 6$ for BNS events [359, 364, 386, 387]. For this, we need the intrinsic merger rate density for the events as a function of redshift. We will assume that the compact binaries originate from stellar populations and adopt the merger rate estimates of Belczynski et

¹Since NSBH merger rates are less certain (see e.g. [44, 385]), we do not consider them here, but we expect general conclusions regarding parameter estimation to largely carry over when signal durations are similar.

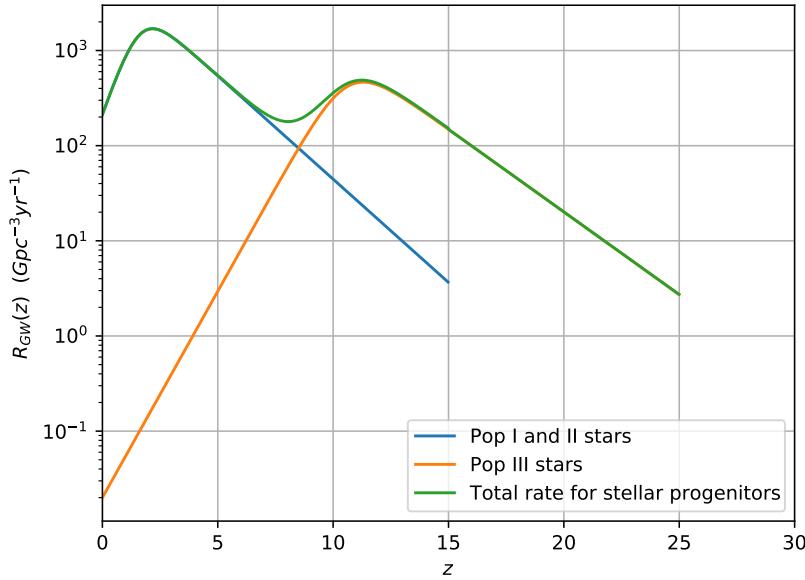


Figure 16.1: The BBH merger rate density according to Oguri’s fit [389] for population I, II, and III stars, as well as the total rate when all the star populations are included.

al. [388] with Oguri’s analytical fit [389]², whose general expression is

$$R_{\text{GW}}(z) = \frac{a_1 e^{a_2 z}}{e^{a_3 z} + a_4}. \quad (16.1)$$

Here R_{GW} is expressed in $\text{Gpc}^{-3} \text{yr}^{-1}$, and the coefficients a_i , $i = 1, \dots, 4$, depend on the considered star populations; see Fig. 16.1. For our purpose, we consider the combination of population I and II stars for BNS, and populations I, II, and III for BBH, as the contribution of the latter type of stars is important only at redshifts of $\gtrsim 4$. However, these relations are rescaled to match the local merger rate estimates obtained observationally by LIGO and Virgo so far; see [326]. In this work, we focus on the lowest, the median, and the highest local rate for each type of event. For BNS, the lowest, median, and highest local rates are, respectively, $80 \text{ Gpc}^{-3} \text{yr}^{-1}$, $320 \text{ Gpc}^{-3} \text{yr}^{-1}$, and $810 \text{ Gpc}^{-3} \text{yr}^{-1}$, which are obtained by changing the value of a_1 to 2480, 9920, and 25110, respectively. On the other hand, for the BBH events, we apply a multiplicative constant to the sum of the population I and II and the population III rates, equal to 0.0709, 0.112, and 0.178 for the lowest, median, and highest local rates, which are $15.1 \text{ Gpc}^{-3} \text{yr}^{-1}$, $23.8 \text{ Gpc}^{-3} \text{yr}^{-1}$, and $37.9 \text{ Gpc}^{-3} \text{yr}^{-1}$, respectively.

An intrinsic merger rate density $R_{\text{GW}}(z)$ is then converted to an *observed* merger rate

²Strictly speaking, this merger rate distribution refers to BBH mergers. However, when computing the merger rate density (see e.g. [386–388]), one assumes a time delay distribution (e.g. $P(t_d) \propto 1/t_d$), with a minimum time delay that is higher for BBH than for BNS. Using the distribution of [388] for both BNS and BBH (with some overall rescaling) then implies that we will underestimate the BNS merger rate density [387] and hence the frequency of overlaps involving BNS signals.

density as a function of redshift by multiplying by the differential comoving volume [359]:

$$R_{\text{GW}}^{\text{obs}}(z) = R_{\text{GW}}(z) \frac{dV_c}{dz}(z). \quad (16.2)$$

To obtain dV_c/dz , we assume the *Planck13* cosmological model [390] of ASTROPY [391, 392].

As a next step, we simulate the population of systems by constructing a “catalog”, and determine which events are detected. For the BBH population, we assume that the masses follow the “*power law + peak*” distribution presented in Ref. [326] for the primary component mass, and the corresponding power law distribution for the mass ratio, through which we sample the secondary mass [326]. For BNS events, we distribute component masses uniformly, where for the primary mass $m_1 \in [1, 2.5] M_\odot$, and the secondary mass $m_2 \in [1 M_\odot, m_1]$. Events are distributed over comoving distance D according to $R_{\text{GW}}(z)$, converting between D and z using the above-mentioned cosmology and cutting off at the maximum redshifts z_{max} stated above. Sky positions and unit normals to the orbital plane are taken to be uniform on the sphere.

In this work, we assume a network of two CEs located at the LIGO Hanford and Livingston sites and one ET located at the Virgo site. We calculate the optimal SNRs in the three observatories for each event. They are added in quadrature to obtain a network SNR. In computing SNRs, we only consider the inspiral part of binary coalescence so that in the SPA (see Sec. 3.1) [72] and for a single interferometer [393]

$$\text{SNR} = \frac{1}{2} \sqrt{\frac{5}{6}} \frac{1}{\pi^{2/3}} \frac{c}{D(1+z)^{1/6}} \left(\frac{G\mathcal{M}_c}{c^3} \right)^{5/6} g(\theta, \phi, \psi, \iota) \sqrt{I(M)}. \quad (16.3)$$

Here, \mathcal{M}_c is the chirp mass in the source frame (2.50). The geometric factor is given by

$$g(\theta, \phi, \psi, \iota) = \left(F_+^2(\theta, \phi, \psi) (1 + \cos(\iota)^2)^2 + 4F_\times^2(\theta, \phi, \psi) \cos(\iota)^2 \right)^{1/2}, \quad (16.4)$$

where $F_{+,\times}$ are the beam pattern functions (see Sec. 2.2.4) in terms of sky position (θ, ϕ) and polarization angle ψ , while ι is the inclination angle. We take the ET to consist of three detectors with 60° opening angle, arranged in a triangle with sides of 10 km [394], and add the corresponding SNRs in quadrature; for CE we assume a single L-shaped detector of 40 km arm length [50]. Finally,

$$I(M) = \int_{f_{\text{low}}}^{f_{\text{high}}} \frac{f^{-7/3}}{S_n(f)} df. \quad (16.5)$$

Here f_{low} is a low-frequency cut-off that depends on the observatory; we set $f_{\text{low}} = 5$ Hz for both ET and CE, though lower values may be achieved in the case of ET [31, 395].

For f_{high} we use the ISCO frequency³:

$$f_{\text{high}}(m_1, m_2, z) = \frac{1}{1+z} \frac{1}{6\pi\sqrt{6}} \frac{c^3}{GM}, \quad (16.6)$$

where $M = m_1 + m_2$ is the total mass. We take the noise PSD $S_n(f)$ to be ET-D in the case of ET [30, 31]; for the projected PSD of CE, we use the one from [50].

The network SNR – denoted SNR_{net} – is defined in Eq. (4.12) and the sum is over the two CE and the one (triangular) ET observatories. We consider an event detectable if the network SNR is above 13.85 ($= \sqrt{3} \times 8$), without imposing SNR thresholds in individual observatories. For the BNS and BBH mass ranges considered here, it means that detection rates will mainly be driven by the CE detectors. Still, we note that ET will have an advantage at higher masses [396].

Finally, signals will be present in a detector for a duration given by

$$\tau = 2.18 \left(\frac{1.21 \text{ M}_\odot}{\mathcal{M}_c} \right)^{5/3} \left[\left(\frac{100 \text{ Hz}}{f_{\text{low}}} \right)^{8/3} - \left(\frac{100 \text{ Hz}}{f_{\text{high}}} \right)^{8/3} \right] \text{ s.} \quad (16.7)$$

Simulated catalogs of events happening over a year are constructed as follows. The year is split into a grid where each cell corresponds to one second. Merger times are drawn from a uniform distribution over these cells. For a given type of event (BNS or BBH), one associates to each merger time a mass pair, redshift, sky position, and orientation of the orbital plane drawn from the corresponding distributions, as well as a signal duration computed from Eq. (16.7). Doing this for the three choices of local merger rate, and in each case putting together the BNSs and BBHs, catalogs of events are obtained. Finally, within each catalog, it is assessed which events will be detectable with the ET-CE network according to the criteria spelled out above, leading to an overview of what we may expect to be contained in one year’s worth of data. In particular, we can check how often and in what way events tend to overlap, depending on their types.

16.1.2 Overlap Estimates

Before imposing detectability thresholds, the three different local merger rates give the following typical numbers of events happening over one year: $\sim 59000, 93000, 148000$ BBH events, and $286000, 1145000, 2900000$ BNS events for the low, median, and high local rates, respectively. The network of two CEs and one ET will detect 93% of BBHs and 35% of BNSs within the considered volumes. The number of detected signals is

³Note that this equation differs from Eq. (2.63) by a factor $1/(1+z)$, added here to account for the expansion of the Universe.

	# detections	SNR _{net}	# SNR _{net} > 250	# SNR _{net} > 100	# SNR _{net} > 50	# SNR _{net} > 20
BBH						
Low rate	53756	81.1 ^{+94.2} _{-57.3}	3069 (5%)	20605 (35%)	40063 (68%)	52239 (89%)
Median rate	85725	81.3 ^{+93.9} _{-57.5}	4972 (5%)	33148 (39%)	63958 (75%)	83333 (97%)
High rate	137225	81.5 ^{+94.2} _{-57.4}	7860 (6%)	53419 (39%)	102766 (75%)	133460 (97%)
BNS						
Low rate	98898	19.2 ^{+22.1} _{-4.9}	17 (0.017%)	298 (0.30%)	2712 (2.7%)	44350 (48%)
Median rate	396793	19.1 ^{+22.0} _{-4.8}	73 (0.018%)	1257 (0.32%)	10659 (2.7%)	177296 (45%)
High rate	1004525	19.1 ^{+22.1} _{-4.8}	196 (0.020%)	3255 (0.32%)	27135 (2.7%)	448610 (45%)

Table 16.1: The number of events detected by a network of two CEs and one ET in one year of simulated data, the median network SNRs and their 90% spreads, and the detection numbers and percentages (in brackets) for different choices of minimum network SNR.

Rate	BBH mergers > 1	BNS mergers > 1	Any mergers > 1
Low rate	48	310	750
Median rate	127	2412	7347
High rate	303	15581	20149

Table 16.2: The number of seconds in a year with at least two mergers occurring, depending on their types.

shown in Table 16.1 for the three local rates, along with median and 90% spreads on SNRs and a breakdown of detections according to their loudness.

Within our simulated event catalogs, we can look at the numbers of detected signals that overlap depending on the types. We focus on two quantities: (i) the number of seconds in a year where at least two detected signals have their merger, and (ii) the typical number of mergers that happen during the time a given signal is in a detector's sensitivity band.

The number of seconds in a year having at least two mergers taking place is given in Table 16.2; clearly, this will happen frequently over a year. Indeed, we find that even more than two mergers can occur within the same second. The proportion of detected signals merging together with at least one other goes up with an increasing local merger rate, potentially reaching thousands per year.

In addition to the scenario where different CBC mergers happen in the same second, we investigate the typical number of mergers happening over the entire duration of a BNS event while it is in band, depending on their type; see Table 16.3 and Fig. 16.2. Because BNS events are in the detector band for a long time (several hours for $f_{\text{low}} = 5$ Hz), quite some such overlaps will indeed occur. If one does the same for BBHs, one finds that either zero or one BBH or BNS merger (at 90% confidence) will happen in its duration; this is due to BBH events being shorter (the median duration being ~ 45 s).

Before moving on to parameter estimation issues, let us briefly look at other future GW observatories that are being planned or considered. Constructing simulated catalogs

Rate	Number of BBH mergers	Number of BNS mergers	Number of any type
Low rate	8^{+10}_{-5}	16^{+16}_{-8}	25^{+23}_{-12}
Median rate	13^{+14}_{-7}	62^{+58}_{-27}	76^{+77}_{-33}
High rate	21^{+21}_{-11}	157^{+144}_{-66}	178^{+164}_{-75}

Table 16.3: Typical numbers of compact binary mergers happening while a BNS signal is in band.

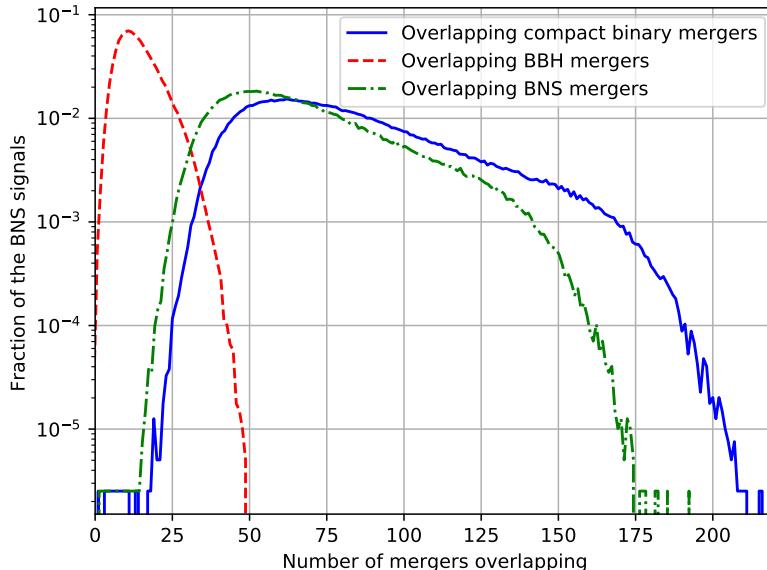


Figure 16.2: Fraction of detected BNS mergers with a given number of compact binary mergers (blue), BBH mergers (red), and BNS mergers (green) taking place while the BNS signal is in band.

of detectable sources in the same way as above and focusing on the high local merger rate, we find that for a year, Advanced LIGO+ [327] will typically have no events merging within the same second, and only a few occurrences of a BBH merging in the duration of a BNS (assuming $f_{\text{low}} = 15$ Hz). For Voyager [397] we find $\mathcal{O}(1)$ instances of two events merging within the same second, and BNS signals will typically have at most one other signal's merger in their duration (for $f_{\text{low}} = 10$ Hz). These numbers refer to signals detectable with a single interferometer (with an SNR threshold 8) rather than with a network of them, but it will be clear that overlapping signals are going to become an important consideration mainly in the 3G era.

Run	BBH-BBH	BBH-BNS	BNS-BNS
Low rate	5	57	416
Median rate	11	304	6752
High rate	15	1594	41306

Table 16.4: Number of pairs of binary coalescence events with both SNRs between 15 and 30, and such that their mergers occur within 2 seconds or less of each other.

16.2 Parameter Estimation Setup

Having established that 3G detectors will see numerous overlapping signals whose mergers occur very close in time, we want to find what this will imply for parameter estimation. To this end, we simulate BBH and BNS signals in a network consisting of one ET and two CE observatories, assuming stationary, Gaussian noise following the PSDs as in the previous section.

Since we expect parameter estimation biases to be more pronounced when SNRs of overlapping signals are similar to each other, and on the low side, we focus on network SNRs roughly between 15 and 30. We consider overlapping events whose merger times coincide (as a proxy for merger within the same second) or are separated by 2 seconds, again because these are the types of scenarios where biases will likely be the largest. The number of overlaps from the previous section satisfying these criteria is given in Table 16.4, for different local merger rates; we see that they will be fairly common.

In our parameter estimation studies, for definiteness, we take the BBH events to have masses similar to those of either GW150914 (a higher-mass, shorter-duration signal) or GW151226 (a lower-mass, longer-duration event). For BNSs, we take the masses to be similar to those of GW170817. Overlapping signals are given different injected sky locations. All analyses are done with three different noise realizations. For each example of overlapping signals, parameter estimation is also done on the individual signals, for the same noise realizations, to assess what biases occur. Fig. 16.3 provides an overview of the various overlap scenarios considered in the rest of this work, in terms of masses and SNRs.

To reduce the computational cost, we focus on non-spinning sources. A BBH signal is then characterized by parameters $\boldsymbol{\theta} = \{m_1, m_2, \alpha, \delta, \iota, \psi, D_L, t_c, \varphi_c\}$, where m_1, m_2 are the component masses, (α, δ) specifies the sky position in terms of right ascension and declination, ι and ψ are respectively the inclination and polarization angles which specify the orientation of the orbital plane with respect to the line of sight, D_L is the luminosity distance, and t_c and φ_c are respectively the time and phase at coalescence. BNS signals have two additional parameters: (Λ_1, Λ_2) , corresponding to the (dimensionless) tidal deformabilities [398–402].

In this work, we focus specifically on potential biases in *intrinsic* parameters. For BBHs, results will be shown for the total mass $M = m_1 + m_2$ and mass ratio $q = m_2/m_1$ (with the convention $m_2 \leq m_1$). For BNSs, we show chirp mass \mathcal{M}_c (2.50) instead of total mass since that parameter is usually the best-determined one for long signals. As the individual tidal deformabilities tend to be poorly measurable for the SNRs considered here, we will be showing results for a parameter $\tilde{\Lambda}$ defined as [403]

$$\tilde{\Lambda} = \frac{16}{13} \sum_{i=1,2} \Lambda_i \frac{m_i^4}{M^4} \left(12 - 11 \frac{m_i}{M} \right), \quad (16.8)$$

since this is how tidal deformabilities enter the waveform phase to leading (5PN) order [398].

The data is analyzed using the usual Bayesian framework, relying on Bayes' theorem (4.17) and the usual GW likelihood (4.18), used for each signal individually. Due to computational limitations, in our parameter estimation studies, we use a lower frequency cut-off of $f_{\text{low}} = 23$ Hz. Since both ET and CE will be sensitive down to lower frequencies than that, we expect our choice will lead to conservative estimates of parameter estimation biases, as the same signal will, in reality, accumulate more SNR when visible in the detector already from a lower frequency.

Our choices for the *prior probability density* $p(\boldsymbol{\theta}|\mathcal{H}_{\text{GW}})$ in Eq. (4.17) are similar to what has been used for the analyses of real data when BBH or BNS signals with masses similar to the ones specified in Fig. 16.3 were present. In all cases, we sample uniformly in component masses. For the GW150914-like signals, we do this in the range $m_1, m_2 \in [10, 80] M_{\odot}$. For analyzing the GW151226-like signals, the component mass range is $m_1, m_2 \in [3, 54.4] M_{\odot}$, and in addition we restrict chirp mass to $\mathcal{M}_c \in [5, 20] M_{\odot}$ and mass ratio q to the range $[0.05, 1]$. For BNSs, we sample component masses in the range $m_1, m_2 \in [1, 2] M_{\odot}$, restricting $\mathcal{M}_c \in [0.7, 2] M_{\odot}$, while tidal deformabilities are sampled uniformly in the range $\Lambda_1, \Lambda_2 \in [0, 5000]$. When we show PDFs for the derived quantity $\tilde{\Lambda}$, they will have been reweighted with the prior probability distribution of this parameter induced by the flat priors on component masses and Λ_1, Λ_2 , such as to effectively have a uniform prior on $\tilde{\Lambda}$.

To sample the likelihood function in Eq. (4.18), we use the LALINFERENCE library [148], and specifically the LALINFERENCE_MCMC algorithm. The waveforms we use for the BNS and BBH signals are *IMRPhenomD_NRTidalv2* [110, 111, 404] and *IMRPhenomD* [103, 104] respectively, both computed with the waveform library LALSIMULATION. To inject the signals and add noise to them, we use standard tools available within the LALSIMULATION package. All these codes are openly accessible in LALSUITE [405].

Before performing parameter estimation, we verify the detectability of the individual signals in the overlap scenarios of Fig. 16.3 using the PyCBC software package [321]. We inject overlapping signals in noise generated from the PSD and check that the individual signals show up as triggers with masses consistent between detectors at a network

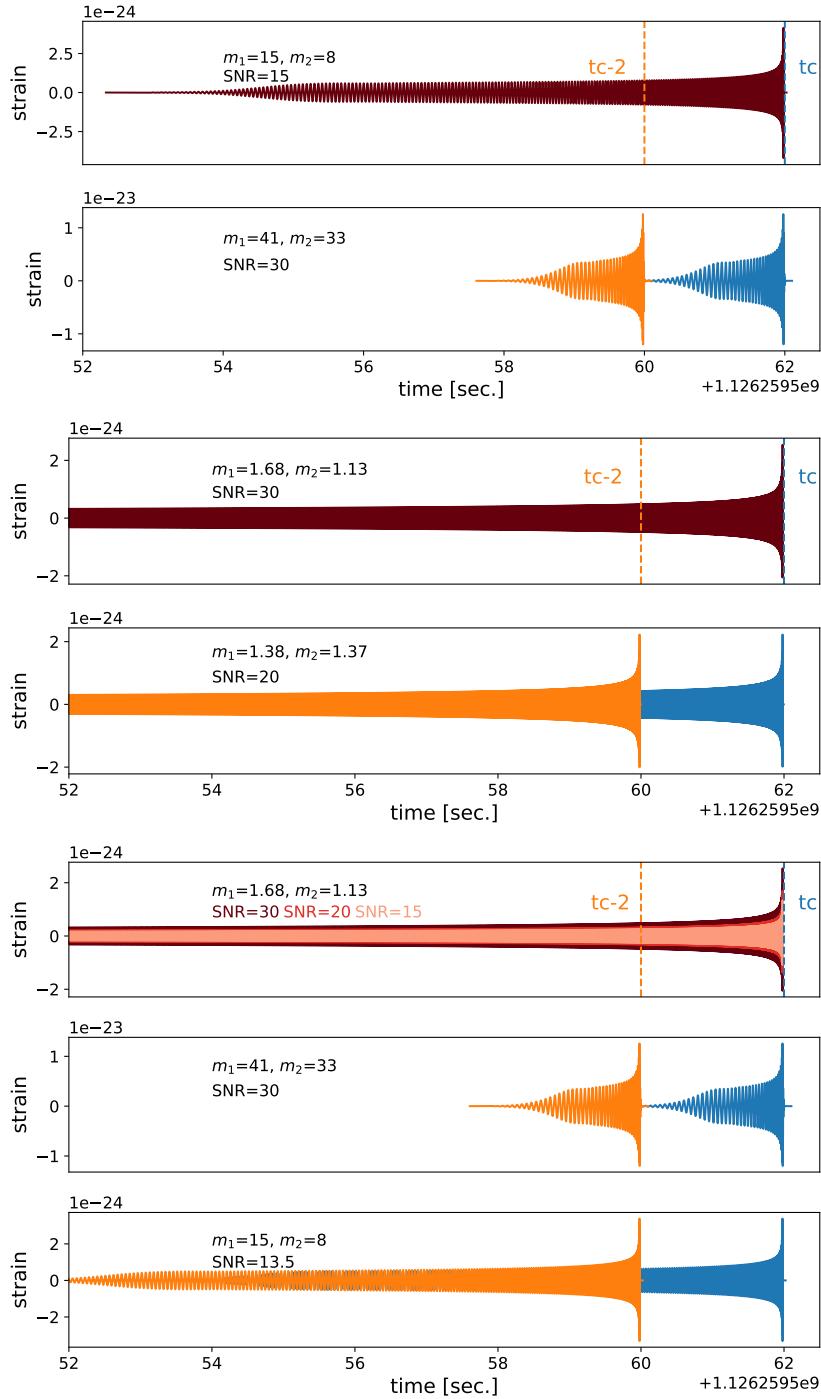


Figure 16.3: Individual waveforms and the overlap scenarios considered in our simulations. All signals are injected in 3 different simulated noise realizations for a 3G detector network. Signals are either overlapped using the same end time (blue waveforms) or 2 seconds earlier than the “primary” signal’s end time (orange waveforms). *Top two panels:* Overlapping waveforms in the case of two BBH signals. The higher-mass BBH signal (bottom; GW150914-like) is overlapped with the lower-mass BBH signal (top; GW151226-like). *Central two panels:* Overlapping waveforms in the case of two BNS signals. *Bottom three panels:* BNS signals (top) with an SNR of 30, 20, or 15 being overlapped with either a high-mass BBH signal (middle; GW150914-like) or a low-mass BBH signal (bottom; GW151226-like).

BBH overlapped	BNS (SNR = 30)			BNS (SNR = 20)			BNS (SNR = 15)		
	\mathcal{M}	q	$\tilde{\Lambda}$	\mathcal{M}	q	$\tilde{\Lambda}$	\mathcal{M}	q	$\tilde{\Lambda}$
GW150914- tc	0.0112	0.00915	0.0277	0.0162	0.0204	0.0275	0.0297	0.0323	0.00947
GW150914- tc-2	0.0320	0.0389	0.0168	0.0235	0.0273	0.0331	0.0704	0.0840	0.0218
GW151226- tc	0.00754	0.00748	0.0113	0.0123	0.0139	0.0173	0.0403	0.0516	0.0305
GW151226- tc-2	0.0187	0.0220	0.0309	0.0227	0.0233	0.0259	0.0521	0.0513	0.0159

Table 16.5: Values of the KS statistic comparing PDFs for BNS parameters (columns) in the BNS+BBH overlap scenarios (rows) with the corresponding PDFs when there is no overlapping BBH signal. The small numbers indicate the absence of significant bias. The numbers shown here correspond to the PDFs in the top panel of Fig. 16.4.

SNR above a threshold of 8. This turns out to be true for all the cases considered, except for two BBH signals merging at the same time. In the latter case, we still have triggers in individual detectors, but with masses differing by up to $\sim 5 M_{\odot}$. Using the SNRs in single detectors as detection statistics, detection is still achieved. For all scenarios, the end times of individual signals tend to be identified with a precision of a few milliseconds [406]; when subsequently performing parameter estimation, we use a prior range for end time that is centered on the true end time, leaving an interval of 0.1 s on either side.

For parameter estimation, all simulations are done with three different noise realizations. Sec. 16.4 gives the same results for the two other noise realizations. Our conclusions are stable over noise realizations.

As usual, the one-dimensional PDF $p(\lambda|\mathcal{H}_{\text{GW}}, d)$ for a particular parameter λ is obtained from the joint PDF $p(\boldsymbol{\theta}|\mathcal{H}_{\text{GW}}, d)$ by integrating out all other parameters (see Sec. 4.2.2). In assessing the effect on parameter estimation of overlapping signals in various ways, we will frequently be comparing one-dimensional PDFs for the same parameter in different situations. A convenient way of quantifying the difference between two distributions $p_1(\lambda)$ and $p_2(\lambda)$ is through the *Kolmogorov-Smirnov (KS) statistic* [407, 408]. Let $P_1(\lambda)$, $P_2(\lambda)$ be the associated *cumulative* distributions; then the KS statistic is just the largest distance between these two:

$$\text{KS} = \sup_{\lambda} |P_1(\lambda) - P_2(\lambda)|. \quad (16.9)$$

By construction, this yields a number between 0 and 1; if the KS statistic is close to zero, then the distributions $p_1(\lambda)$ and $p_2(\lambda)$ will be considered close to each other.

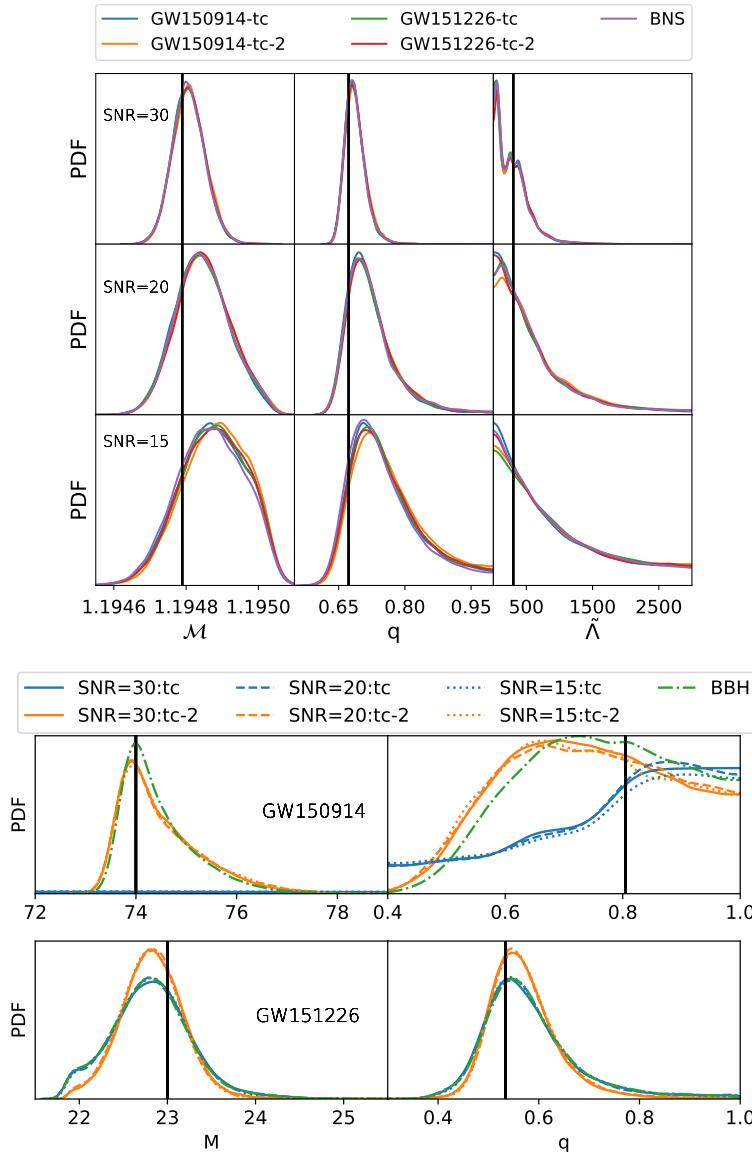


Figure 16.4: *Top:* Posterior PDFs showing estimation of intrinsic parameters when the BNS signal has SNR = 30 (top row), SNR = 20 (middle row), and SNR = 15 (bottom row). Results are shown for the cases when the GW150914-like signal ends at the same time as the BNS signal (GW150914-tc), when it ends 2 seconds earlier (GW150914-tc-2), when the GW151226-like signal ends at the same time as the BNS (GW151226-tc), when it ends 2 seconds earlier (GW151226-tc-2), and finally when the injected signal is only the BNS (BNS). The true values of the parameters are indicated by vertical black lines. We do not observe any bias in the recovered posteriors. *Bottom:* Posterior PDFs for total mass and mass ratio, for the GW150914-like signal (top panel) and the GW151226-like signal (bottom panel) when they are respectively being overlapped with a BNS signal of SNR = 30 (solid lines), SNR = 20 (dashed lines), and SNR = 15 (dotted lines). The overlaps are being done when the BBH and the BNS end at the same time (tc), and when the BBH ends 2 seconds before the BNS (tc-2). Finally, posterior PDFs for the two BBH signals without overlap are shown as green, dashed-dotted lines (BBH). The injected parameter values are indicated by black vertical lines. When the GW150914-like signal ends at the same time as the BNS, the event is not found by the sampler. For the GW151226-like signal, there is a slight deterioration in the recovery but we retain measurability. When the BBHs merge two seconds before the BNS, we do not observe bias.

BNS overlapped	GW150914- tc		GW150914- tc-2		GW151226- tc		GW151226- tc-2	
	<i>M</i>	<i>q</i>	<i>M</i>	<i>q</i>	<i>M</i>	<i>q</i>	<i>M</i>	<i>q</i>
BNS (SNR = 15)	–	–	0.0504	0.0807	0.00933	0.0117	0.0687	0.0657
BNS (SNR = 20)	–	–	0.0427	0.0698	0.0107	0.0106	0.0727	0.0700
BNS (SNR = 30)	–	–	0.0379	0.0673	0.0187	0.183	0.0819	0.0793

Table 16.6: Values of the KS statistic comparing PDFs for BBH parameters (columns) in the BNS+BBH overlap scenarios (rows) with the corresponding PDFs when there is no overlapping BNS signal. In the case of a GW150914-like signal merging at the same time as a BNS, the sampler fails to find the signal, but other scenarios are not so problematic. For GW151226, the slightly higher values for the **tc-2** case compared to the **tc** case are likely due to the signals placed in a slightly different part of the noise stream (two seconds earlier) from the BBH-only cases used for comparison. The numbers shown here correspond to the PDFs in the bottom panel of Fig. 16.4.

16.3 Results

16.3.1 Overlap of a BNS Signal with a BBH Signal

First, we look at the results of parameter estimation for the overlap of a BNS signal with a BBH, either ending at the same time or with the BBH signal ending 2 seconds earlier than the BNS. This scenario is shown in the three bottom panels of Fig. 16.3. We perform parameter estimation first on the BNS and then on the BBH, with priors as specified in the previous section.

16.3.1.1 BNS Recovery

The top panel of Fig. 16.4 shows posterior probability distributions for intrinsic parameters characterizing the BNS signal for three different SNRs of the BNS and the different overlap scenarios. The PDFs tend to widen with decreasing SNR, as expected. We see that the estimation of the mass parameters is essentially unaffected, regardless of the type of overlapping BBH signal (GW150914-like or GW151226-like) or of its merger time (identical to that of the BNS or 2 seconds earlier). For a given SNR of the BNS, the PDFs for the tidal parameter $\tilde{\Lambda}$ differ slightly more between the overlap scenarios. However, we note that most of the information on tides enters the signal at high frequencies, where the detectors are less sensitive; in fact, as shown in Sec. 16.4 (Fig. 16.7), differences in the underlying noise realization tend to have a larger effect on

the measurement of $\tilde{\Lambda}$ than overlapping signals.

We conclude that an overlapping BBH signal does not impact the estimation of the BNS parameters much, even if the BBH merger time is arbitrarily close to that of the BNS. This is corroborated by the KS statistics in Table 16.5, comparing PDFs for the various overlap scenarios with the corresponding PDFs in the absence of overlapping signals. It is reasonable to assume that placing a BBH signal even earlier in the BNS would also have had little impact.

16.3.1.2 BBH Recovery

The bottom panel of Fig. 16.4 shows parameter estimation on the BBHs when the SNR of the BNS signal is varied from 30 to 20 to 15. Table 16.6 has the corresponding KS statistics comparing overlapped signals with PDFs obtained in the absence of overlap. Again, results are shown for a particular noise realization; see Fig. 16.8 in Sec. 16.4 for two other noise realizations. When the BBH signal has its coalescence time 2 seconds earlier than the BNS ($tc-2$ in the figure), the signal is well recovered. However, when the BBH signal and the BNS signal end at the same instant of time, the BBH recovery deteriorates, and in the case of the GW150914-like signal, the sampling process fails to find the signal. For the GW151226-like signal, while the estimates are offset from their true values, there is some measurability of the signal when the times of coalescence of the BBH and BNS are the same. The different outcomes between the GW150914-like and GW151226-like injections are likely due to the short duration of the GW150914-like signal, effectively leading to a distortion of the entire signal when the merger happens at the same instant as the BNS merger. By contrast, the much longer inspiral of the GW151226-like signal implies many more wave cycles for the parameter estimation algorithm to latch on to. Finally, as the SNR over the underlying BNS signal is varied (keeping the SNR of the BBH signal the same), the PDFs for the BBH show essentially no change. Placing a BBH signal only 2 seconds before the BNS merger causes the BBH to be recovered without appreciable biases, so it is reasonable to assume that placing a BBH signal still earlier in the BNS inspiral would also have little effect on its recovery.

16.3.2 Overlap of Two BBH Signals

The scenario analyzed here is the one in the two top panels of Fig. 16.3. Fig. 16.5 shows the posterior PDFs on total mass M and mass ratio q when two BBH signals of different masses are overlapped, compared with parameter estimation on the same signals in situations where there is no overlap (BBH). The corresponding KS statistic values are given in Table 16.7. We find the results to be consistent with statistical fluctuations. Here too, the signals are overlapped once with identical coalescence times (tc), and once

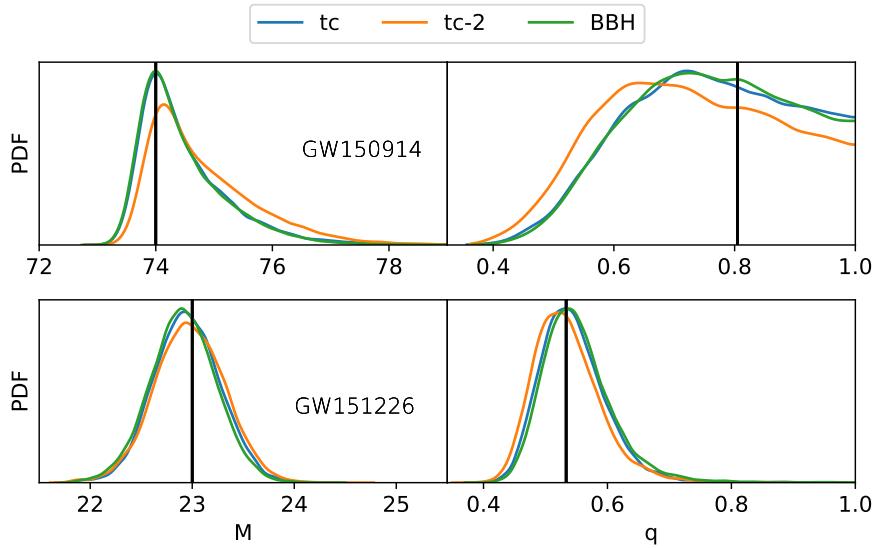


Figure 16.5: Posterior PDFs for total mass and mass ratio when a GW150914-like signal and a GW151226-like signal are being overlapped at the same trigger times (tc) and when the trigger time of the GW150914-like BBH ends 2 seconds earlier ($tc-2$), compared with parameter estimation in the absence of overlap (BBH). The top panel shows the recovery of the GW150914-like signal, and the bottom one that of the GW151226-like signal. Black vertical lines indicate the true values of the parameters. No major bias is observed and variations are within statistical fluctuation.

with one of the signals, GW150914, ending 2 seconds earlier ($tc-2$). The SNRs of the two signals, GW150914-like, and GW151226-like, are 30 and 15, respectively. As can be seen in Fig. 16.5, the two BBH signals' parameters can be extracted without any biases even when they end simultaneously. Again, see Sec. 16.4 for other noise realizations with the same conclusion.

16.3.3 Overlap of Two BNS Signals

Finally, we analyze the simulations in the two middle panels of Fig. 16.3. Fig. 16.6 shows the recovery of BNS parameters for each BNS signal when two BNS signals are being overlapped, again with either the same coalescence times and when one of the BNSs (henceforth BNS2) ends 2 seconds earlier than the other BNS signal (henceforth BNS1). For KS statistic values comparing PDFs with the corresponding non-overlapping cases, see Table 16.8. BNS1 and BNS2, respectively, have SNRs of 30 and 20, and component masses $(m_1, m_2) = (1.68, 1.13) M_\odot$ and $(m_1, m_2) = (1.38, 1.37) M_\odot$. These particular choices cause both signals to have very similar chirp masses. Given these masses, their tidal deformabilities, $\tilde{\Lambda} = 303$ for BNS1 and $\tilde{\Lambda} = 292$ for BNS2, follow the equation of

GW150914-tc		GW150914-tc-2		GW151226-tc		GW151226-tc-2	
M	q	M	q	M	q	M	q
0.0195	0.0109	0.162	0.103	0.0446	0.0478	0.0844	0.127

Table 16.7: Values of the KS statistic comparing PDFs for BBH parameters in the BBH+BBH overlap scenarios with the corresponding PDFs without an overlapping signal. The slightly higher values for the **tc-2** cases are likely due to the signals being in a slightly different part of the noise stream (two seconds earlier) from the BBH-only cases used for comparison. However, in all cases, there is no significant bias. The numbers shown here correspond to the PDFs in Fig. 16.5.

state APR4 [409]; these were the simulated signals used for investigating systematics in the measurements on GW170817 in Ref. [410].

In Fig. 16.6, the top panel shows the posterior PDFs on chirp mass, mass ratio, and tidal deformability for BNS1 when BNS2 ends at the same time (**tc**) and when BNS2 ends 2 seconds earlier (**tc-2**), together with the case where only BNS1 is present in the data (**BNS**). The bottom panels show the same, but for the recovery of BNS2. When the two signals end at the same time, the parameters characterizing BNS1 are recovered,

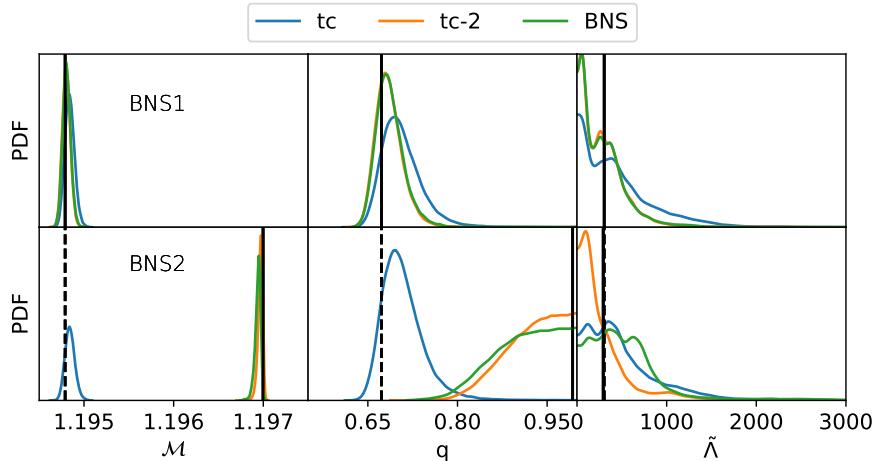


Figure 16.6: Posterior PDFs showing recovery on chirp mass, mass ratio, and tidal deformability $\tilde{\Lambda}$ when two BNSs, referred to as BNS1 and BNS2, are being overlapped at the same time of coalescence (**tc**) and when BNS2 ends 2 seconds earlier than BNS1 (**tc-2**). These are compared with results in the absence of overlap (**BNS**). The top panel is for the recovery of BNS1 and the bottom one is for the recovery of BNS2. The solid black vertical lines indicate the injected values of the source being recovered each time. We note that when the times of coalescence of the two BNSs are the same, the parameter estimates recovered are those of BNS1, whose injected values are also shown in the bottom panel as dashed vertical black lines. Therefore, in this case, the fainter BNS signal is not recovered.

BNS1 (tc)			BNS1 (tc-2)			BNS2 (tc)			BNS2 (tc-2)		
\mathcal{M}	q	$\tilde{\Lambda}$	\mathcal{M}	q	$\tilde{\Lambda}$	\mathcal{M}	q	$\tilde{\Lambda}$	\mathcal{M}	q	$\tilde{\Lambda}$
0.269	0.270	0.202	0.0309	0.0216	0.0129	1.0	0.955	0.0762	0.384	0.0951	0.368

Table 16.8: Values of the KS statistic comparing PDFs for BNS parameters in the BNS+BNS overlap scenarios with the corresponding PDFs without an overlapping signal; see also Fig. 16.6. We see that the numbers are higher for both BNSs when they end at the same time; in fact, the measured parameters for BNS2 are those of BNS1. However, when the BNSs merge 2 seconds apart, the values are much lower, showing that the biases largely disappear. The numbers shown here correspond to the PDFs in Fig. 16.6.

likely because of the higher SNR of BNS1. As the tidal deformabilities of the two sources are so close, the PDFs for $\tilde{\Lambda}$ look similar in all cases. However, also looking at the mass parameters, parameter estimation is rather robust when the signals end 2 seconds apart.

16.4 Impact of the Noise Realizations

We have performed all our simulations in three different noise realizations. To avoid plots getting too busy, in Sec. 16.3 we only showed results for one of these; here we also give them for the other two noise realizations.

In the case of a BNS overlapping with a BBH, the measurements on the BNS are shown in Fig. 16.7 and those on the BBH in Fig. 16.8. The corresponding KS values are given in Tables 16.9 and 16.10, respectively. For measurements of the mass parameters of the BNS, we find that the results are consistent between noise realizations. For the tidal parameter $\tilde{\Lambda}$, the PDFs differ somewhat more; compare the right columns in the two panels of Fig. 16.7. This is likely because most of the information on tides enters the signal at higher frequencies, where the variance of the noise is larger; hence the measurement of $\tilde{\Lambda}$ will be more affected by the noise realization than the mass measurements, especially when SNRs are not high. Indeed, though not shown here explicitly, for a given overlap situation, KS statistics for $\tilde{\Lambda}$ between different noise realizations tend to be significantly larger than within the same noise realization but for different overlaps. For parameter estimation on the BBH, there are differences in the PDFs for the masses when the BBH merger time coincides with that of the BNS but not so much if it occurs 2 seconds earlier.

In the case of two overlapping BBH signals, parameter estimation results are shown in Fig. 16.9, and KS statistics in Table 16.11. The results are quite robust under a change of noise realization.

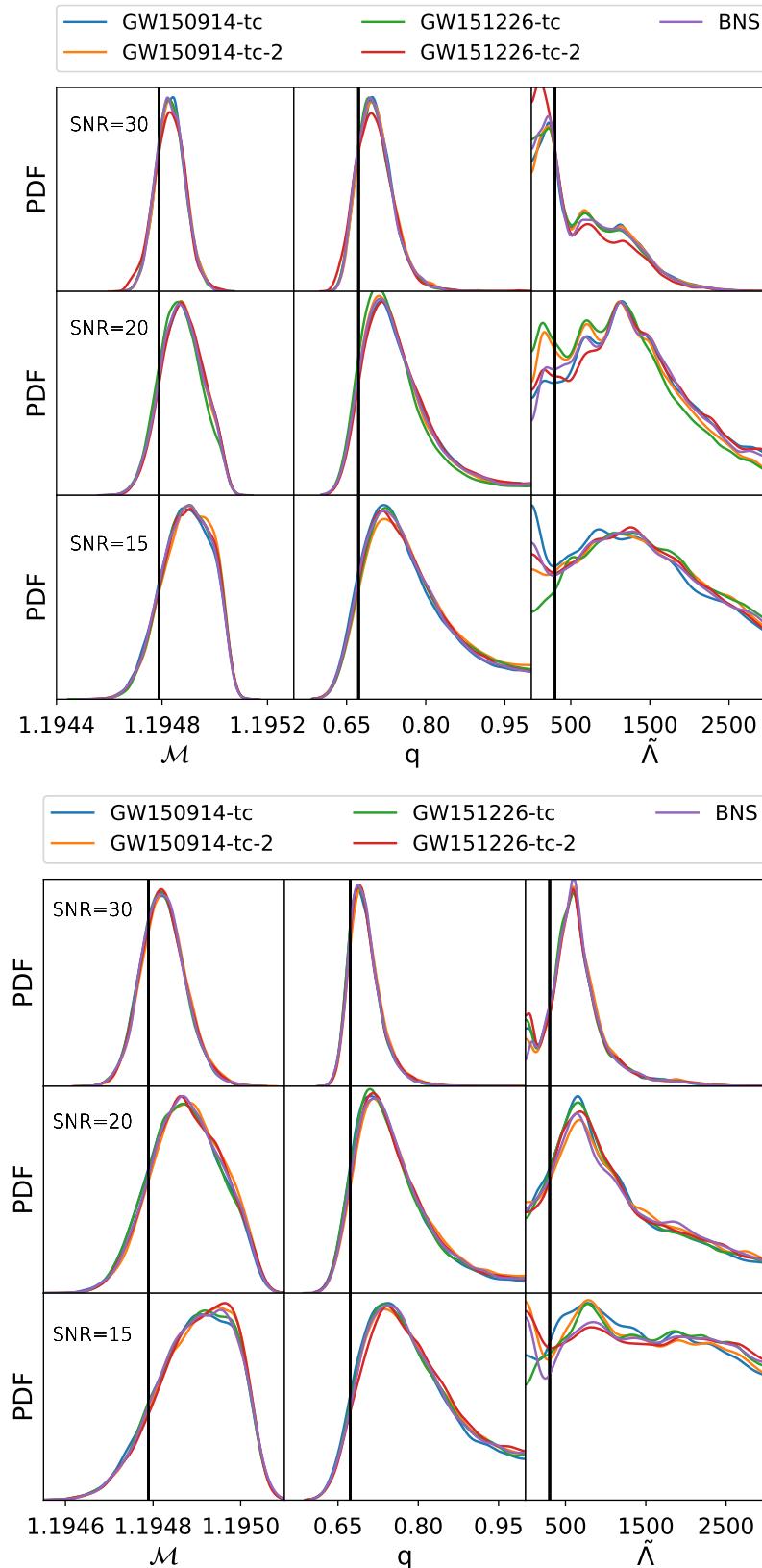


Figure 16.7: Posterior PDFs for BNS parameters when a BNS and BBH signal are being overlapped; same as the top panel of Fig. 16.4 when injections are done in two other noise realizations (left and right panels). The variations for the $\tilde{\Lambda}$ are more affected by the difference in noise realization than by the overlap itself.

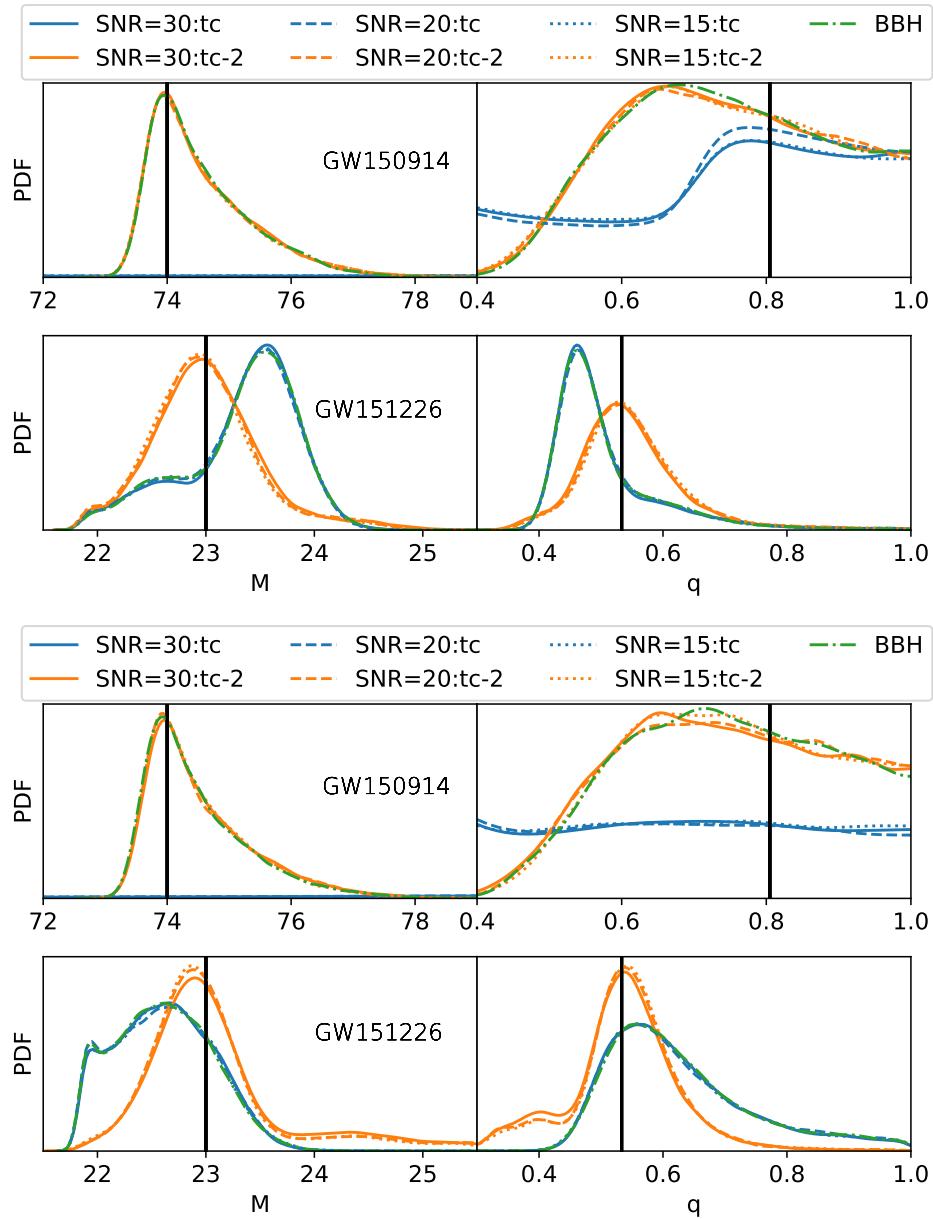


Figure 16.8: Posterior PDFs for BBH parameters when a BNS and BBH signal are overlapped; same as the bottom panel of Fig. 16.4 when injections are done in two other noise realizations.

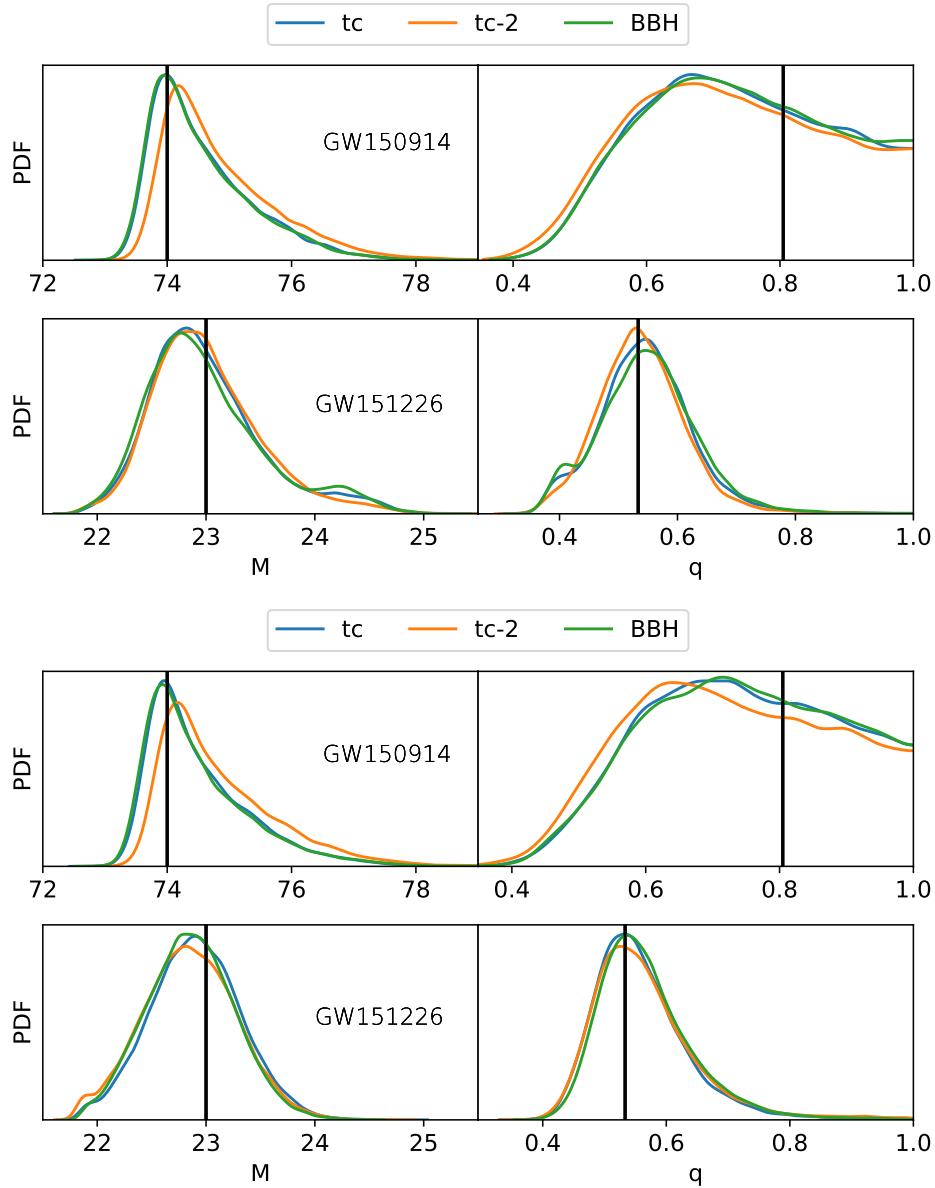


Figure 16.9: Posterior PDFs for BBH parameters when two BBH signals are being overlapped; same as Fig. 16.5 when injections are done in two other noise realizations.

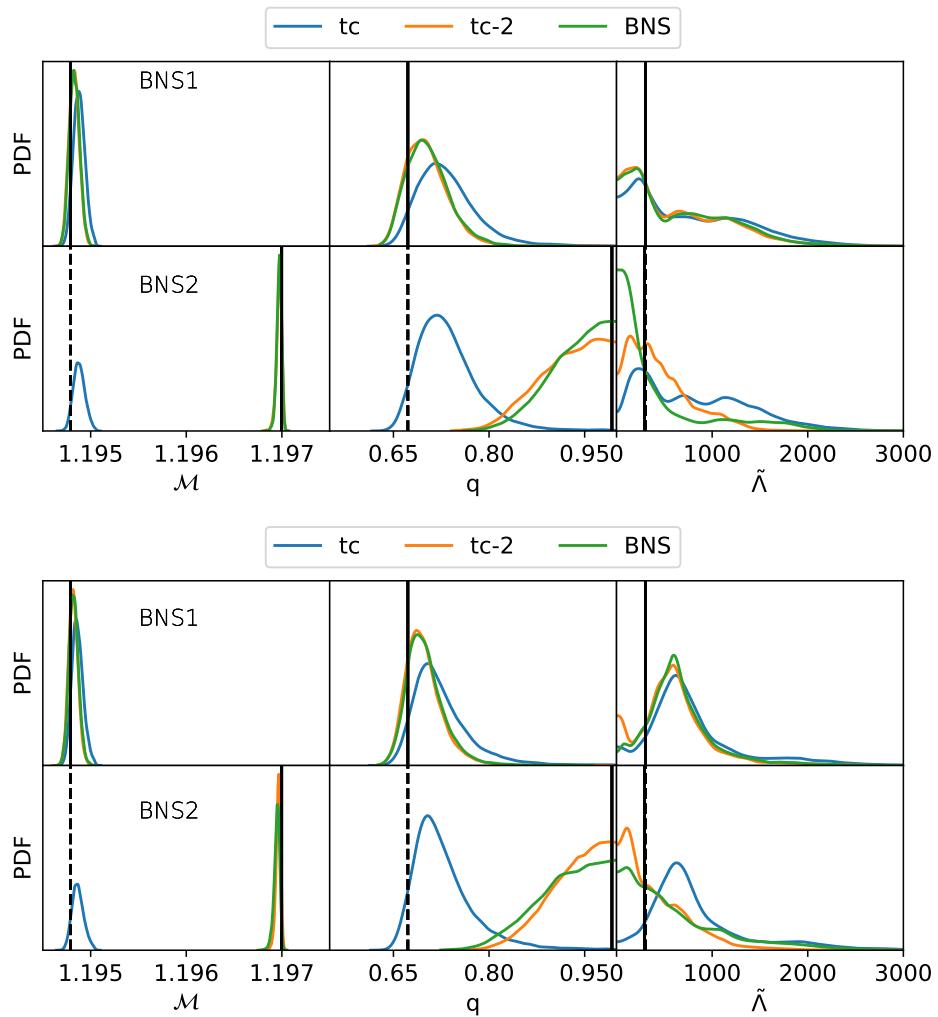


Figure 16.10: Posterior PDFs when two BNS signals are being overlapped; same as Fig. 16.6 when injections are done in two other noise realizations.

BBH overlapped	BNS (SNR = 30)			BNS (SNR = 20)			BNS (SNR = 15)		
	\mathcal{M}	q	$\tilde{\Lambda}$	\mathcal{M}	q	$\tilde{\Lambda}$	\mathcal{M}	q	$\tilde{\Lambda}$
<i>Noise realization 2</i>									
GW150914-tc	0.0267	0.0248	0.0224	0.0106	0.0141	0.0169	0.0146	0.0211	0.0290
GW150914-tc-2	0.0287	0.0282	0.0338	0.00601	0.0108	0.0486	0.0263	0.0308	0.0137
GW151226-tc	0.0125	0.0141	0.0421	0.0376	0.0471	0.0723	0.0155	0.0152	0.0333
GW151226-tc-2	0.0337	0.0346	0.0815	0.0244	0.0258	0.0179	0.0113	0.0108	0.00923
<i>Noise realization 3</i>									
GW150914-tc	0.0140	0.0143	0.0251	0.0236	0.0298	0.0481	0.0114	0.0255	0.0378
GW150914-tc-2	0.0296	0.0396	0.0255	0.0272	0.0218	0.0125	0.0186	0.0125	0.0299
GW151226-tc	0.0135	0.0161	0.0347	0.0215	0.0312	0.0412	0.00750	0.00868	0.0239
GW151226-tc-2	0.0142	0.0140	0.0334	0.0109	0.00833	0.0310	0.0223	0.0292	0.0169

Table 16.9: Values of the KS statistic comparing PDFs for BNS parameters (columns) in the BNS+BBH overlap scenarios (rows) with the corresponding PDFs when there is no overlapping BBH signal when injections are done in two other noise realizations. The numbers shown correspond to the PDFs in Fig. 16.7, noise realization 2, corresponding to the top panel, and noise realization 3 to the bottom panel.

16.5 Conclusions and Outlook

Given regular improvements in the sensitivity of GW detectors and, especially, the planned construction of the next-generation interferometers, it will become increasingly likely that individually detectable GW signals end up overlapping in the data. In this work, we (i) assessed how often different types of overlap will happen in ET and CE, and (ii) tried to quantify the impact this would have on parameter estimation with current data analysis techniques.

To address the question of the nature and frequency of different overlap scenarios, for each of the three possible local merger rates, we constructed a “catalog” of signals in ET and CE, enabling a more in-depth study of overlaps than previous works. We showed that there will be a significant number of signals for which the merger happens within the same second, varying from tens to thousands depending on the local merger rate. Additionally, the substantial increase in the duration of BNS events due to the improved low-frequency sensitivity of 3G observatories will lead to up to tens of other signals overlapping with a given BNS.

Motivated by these results, we performed the first detailed Bayesian analysis study on possible biases that may arise in the future as detection rates become higher and overlapping signals start to occur. We focused on overlapping signals for which the end times were close to each other so that, in particular, there is overlap at times when both signal amplitudes are high; it is in this type of situation that we expect parameter estimation biases to be the most pronounced. Specifically, merger times were taken to

BNS overlapped	GW150914-tc		GW150914-tc-2		GW151226-tc		GW151226-tc-2	
Noise realization 2	M	q	M	q	M	q	M	q
BNS (SNR = 15)	–	–	0.0134	0.011	0.00832	0.00890	0.411	0.398
BNS (SNR = 20)	–	–	0.0104	0.0109	0.0169	0.0172	0.390	0.377
BNS (SNR = 30)	–	–	0.0100	0.0113	0.0140	0.0146	0.367	0.357
Noise realization 3	M	q	M	q	M	q	M	q
BNS (SNR = 15)	–	–	0.0168	0.0100	0.0140	0.0142	0.318	0.131
BNS (SNR = 20)	–	–	0.0189	0.0131	0.0132	0.0137	0.322	0.315
BNS (SNR = 30)	–	–	0.0287	0.295	0.0136	0.0130	0.334	0.327

Table 16.10: Values of the KS statistic comparing PDFs for BBH parameters (columns) in the BNS+BBH overlap scenarios (rows) with the corresponding PDFs when there is no overlapping BNS signal when injections are done in two other noise realizations. As before, when the GW150914-like signal ends at the same time as a BNS, it is not found by the sampling algorithm, but other scenarios are less problematic. The numbers shown correspond to the PDFs in Fig. 16.8, noise realization 2, corresponding to the top panel, and noise realization 3 to the bottom panel.

GW150914-tc		GW150914-tc-2		GW151226-tc		GW151226-tc-2	
M	q	M	q	M	q	M	q
0.0195	0.00854	0.163	0.0395	0.0299	0.0309	0.0417	0.0746
M		M		M		M	
0.0291	0.0110	0.188	0.0625	0.0477	0.0497	0.0225	0.0440

Table 16.11: Values of the KS statistic comparing PDFs for BBH parameters in the BBH+BBH overlap scenarios with the corresponding PDFs for the BBH-only case when injections are done in two other noise realizations. The numbers shown correspond to the PDFs in Fig. 16.9, the upper row corresponding to the upper panel and the lower row to the lower panel.

be either the same (as a proxy for being arbitrarily close to each other), or separated by 2 seconds. Our preliminary conclusions (based on a limited number of investigations) are as follows:

- When BBH signals are overlapping with a BNS signal of similar SNR, parameter estimation on the BNS is hardly affected, even with the merger time of the BBH arbitrarily close to that of the BNS. Presumably, this is due to the much larger number of BNS wave cycles in band compared to the BBH.
- However, in the same scenario, parameter estimation on the BBH can be subject to significant biases if the BBH is high-mass so that its signal is short. That said, the problem largely disappears when the BNS and BBH merger times are separated by 2 seconds or when the BBH has a low mass.

BNS1 (tc)			BNS1 (tc-2)			BNS2 (tc)			BNS2 (tc-2)		
\mathcal{M}	q	$\tilde{\Lambda}$	\mathcal{M}	q	$\tilde{\Lambda}$	\mathcal{M}	q	$\tilde{\Lambda}$	\mathcal{M}	q	$\tilde{\Lambda}$
0.316	0.282	0.0743	0.0385	0.0339	0.0325	1.0	0.936	0.382	0.0271	0.0858	0.248
BNS1 (tc)			BNS1 (tc-2)			BNS2 (tc)			BNS2 (tc-2)		
\mathcal{M}	q	$\tilde{\Lambda}$	\mathcal{M}	q	$\tilde{\Lambda}$	\mathcal{M}	q	$\tilde{\Lambda}$	\mathcal{M}	q	$\tilde{\Lambda}$
0.278	0.257	0.123	0.0475	0.0381	0.0630	1.0	0.902	0.341	0.226	0.101	0.128

Table 16.12: Values of the KS statistic comparing PDFs for BNS parameters in the BNS+BNS overlap scenarios with the corresponding PDFs for the BNS-only case, when injections are done in two other noise realizations. The numbers shown correspond to the PDFs in Fig. 16.10, the upper row corresponding to the upper panel and the lower row to the lower panel.

- When two BBHs with sufficiently dissimilar masses overlap with close-by merger times, parameter estimation on either of the signals will not be much affected.
- When two BNS signals overlap with close-by merger times, parameter estimation will recover the louder signal reasonably well. With a 2-second separation of merger times, good-quality parameter estimation can already be done on the two signals separately.

These results suggest that current parameter estimation techniques will, in several types of situations of interest, already perform reasonably well in the 3G era when applied to overlapping signals, even when the individual signals have similar SNRs, and even when the SNRs are on the low side given the projected distribution for these observatories. Nevertheless, several questions remain. What happens when SNRs are gradually increased? Related to this is the choice of the lower cut-off frequency; to what extent will parameter estimation improve as one goes to $f_{\text{low}} = 5$ Hz or even lower so that signals have a much larger number of wave cycles in the detector’s sensitive band? Though not the focus here, at higher SNRs the use of currently available waveform approximants to analyze BNS signals in 3G detectors would lead to biases in the estimation of $\tilde{\Lambda}$ even in the absence of overlap [411], also motivating further research in waveform modeling. Spins were not included in our study, but it would be of interest to see their effect: large precessing spins will complicate parameter estimation in the case of BBHs, while for BNSs, having access to the spin-induced quadrupole moment can aid in determining tidal deformabilities [412]. Finally, what happens when overlaps involve (much) more than two signals, e.g. long BNS signal overlapping with numerous BBH signals? These questions are left for future work.

To make optimal scientific use of the capabilities of 3G detectors, it will be appropriate to develop Bayesian parameter estimation techniques for which the likelihood function assumes multiple signals to be present in a given stretch of data, e.g. replacing Eq. (4.18) by

$$p(d|\{\boldsymbol{\theta}_i\}, \mathcal{H}_{\text{GW}}) \propto \exp \left[-\frac{1}{2} \left\langle d - \sum_{i=1}^N h(\boldsymbol{\theta}_i) \middle| d - \sum_{i=1}^N h(\boldsymbol{\theta}_i) \right\rangle \right], \quad (16.10)$$

with N the number of signals found by a detection pipeline, and $\boldsymbol{\theta}_i$, $i = 1, \dots, N$, the associated parameters. Additionally, one could let N itself be a parameter to be sampled over, thus allowing for an a priori unknown number of signals in the given stretch of data. In all this, it may be possible to borrow from techniques developed in the context of somewhat related problems in GW data analysis, such as the characterization of a large number of (in this case near-monochromatic) signals from galactic white dwarf binaries in the space-based LISA [413–419], BNSs in the Big Bang Observatory (BBO) [374], or supermassive BH binaries in pulsar timing searches [420].

17

PARAMETER ESTIMATION METHODS FOR ANALYZING OVERLAPPING GRAVITATIONAL-WAVE SIGNALS IN THE THIRD-GENERATION DETECTOR ERA

Chapter 16 and other works [319, 360–363] have established the impact of overlapping signals on data analysis when recovering one of the two signals and neglecting the other. These works employed different techniques, but all have the same conclusions: biases can occur in various scenarios, are most likely when the signals merge close to each other, and depend on the signals' nature.

In Ref. [360], using a Fisher matrix approach, the authors show BNSs are less correlated in overlapping signals. Therefore, their bias becomes important only for close merger times (< 0.1 s). Conversely, the correlation between BBHs is more important, meaning that the bias can happen for larger differences between the merger times. They then proceed to do parameter estimation for overlapping BBHs in a LIGO-Virgo network varying some of the parameters, showing the appearance of biases for merger times close to each other. In [363], the authors use Fisher matrices to study the bias that can occur in the parameters for both overlapping BBHs and BNSs, also finding that the bias becomes more important for short differences in the time of arrival. In [361], the authors focus on a LIGO Voyager scenario, looking at the biases as a function of other parameters than the coalescence time, such as the sky location. For overlapping BBHs, they find more important biases for closer merger times. However, they show that the observed bias for a given difference in merger time can change substantially depending on the sky location of the two events. Moreover, two overlapping BBHs can be mistaken for one strongly precessing BBH. Also, the authors suggest no major bias will occur for overlapping BNS and BBH signals due to the different durations of the two signals. This is corroborated by the analyses done in Chapter 16, where three scenarios are analyzed: two overlapping BBH signals, two overlapping BNS signals, and the overlap of a BBH of varying masses with a BNS. In the latter case, the authors find that there is hardly any effect on the BNS parameter estimation, probably due to the difference in the number of cycles present in-band for this signal. However, in this scenario, the BBH can be affected by significant bias, especially when the BBH has high component masses. The bias mostly disappears when the merger times are separated by more than two seconds.

For two overlapping BBHs, if the total masses, hence the durations, are different, the parameter estimation is done relatively well. However, others [360, 361, 363] have shown that if the two BBHs have similar source properties, biases can be present. Finally, for two overlapping BNSs, it appears that the signal with the highest SNR is relatively well recovered in all cases but not necessarily the quieter signal.

One problem not covered by previous studies is the effect of confusion noise on parameter estimation. Indeed, the high event rate and increased event duration will make for very few periods without signal in-band for the 3G detectors [319]. As a consequence, it will be very difficult to estimate the noise present in the detectors, and additional biases can occur due to a mismodeling of the noise [362]¹. A demonstration of this noise's effect on matched filtering and how the PSD could be computed are presented in [421].

The presence of biases when signals merge close to each other and the relative occurrence of such scenarios based on the estimated rates shows that parameter estimation methods will have to be adapted to be suited for the 3G cases. Indeed, biases in the parameters estimated for the CBCs would impact any direct science case for the CBCs (such as measuring their mass distribution and rate or testing GR [422]), and also some indirectly related ones, such as the search for primordial BHs since it requires subtracting foreground sources [377–379, 423–425].

In this chapter of the thesis and Ref. [426], we look at two possible methods to analyze overlapping signals. One is *Hierarchical subtraction* (HS), where we analyze one signal (typically the loudest), then subtract the maximum-likelihood template before studying the second one. However, if an important bias happens when analyzing the first signal, the parameters of both events might be biased². Optionally, one can also perform a third run, subtracting the maximum-likelihood parameters for the second event and re-analyzing the dominant signal to reduce the bias in its recovery. Still, this is not guaranteed to lead to unbiased results. Therefore, we also implement a *joint parameter estimation* (JPE) framework, where the two signals are analyzed simultaneously to account for the entire model. In principle, this should be the most complete model one can use. Due to the high dimensionality of the parameter space, combined with the large duration of the signals in the 3G detectors, this framework is substantially slower than HS³. It would be nearly impossible to follow the predicted rates for an ET and CE network using a simple JPE like the one used in this work. Such constraints could be alleviated by using recently developed techniques, like relative binning [354–356],

¹We will not consider this in this study, and we will only look at overlapping binary signals.

²In this chapter, the term “bias” is used colloquially and denotes any change in the recovered posteriors due to the other signal's presence.

³For our experiments, the two frameworks were run on the same cluster using 16 *Intel(R) Xeon(R) Gold 6152* CPUs. The average run time for JPE is 23.8 days, while for HS, the first run took an average of 6.3 days, the second run an average of 4.3 days, and the last run took an average run time of 6.1 days. So, on average, JPE takes 7 more days to complete than HS if we perform the three runs. If one is satisfied with the two first runs, the difference between the two approaches goes up to about two weeks.

adaptive frequency resolution [353], or ML [297, 299], see Chapter 18 for a proof-of-concept use of ML for parameter inference for overlapping signals. One could also count on the development of more powerful computational methods, such as quantum computing [358], but it is difficult to have an idea of the state of such methods by the time the 3G detectors get online. Still, it is important to start preparing for the future of 3G detectors now. Therefore, it is crucial to start looking at the parameter estimation of overlapping signals to have the bases to build on. Due to the limited computational resources, this work focuses on the parameter estimation of two overlapping BBH signals.

17.1 Description of the Methods

When performing GW data analysis on CBC signals, our objective is to find the posterior probability density function (PDF) of the binaries' parameters ($\boldsymbol{\theta}$): $p(\boldsymbol{\theta}|d, \mathcal{H})$, where d represents the data, and \mathcal{H} is the hypothesis under which we work (*e.g.* \mathcal{H}_{GW} when there is a GW signal in the data). This can be done using Bayes' theorem (4.17) as explained in Sec. 4.2.

For GW inference, $d(t)$ is the output of the interferometers, which can be seen as made of a noise component $n(t)$ and, under the signal hypothesis, a GW component $h(t)$:

$$d(t) = n(t) + h(t). \quad (17.1)$$

In our scenario, the GW component can consist of one or more signals. In the latter case, $h(t) = \sum_{i=1}^N h_i(t)$, where $h_i(t)$ is the representation of each GW signal, and N is the total number of GW signals present in the data stretch under consideration.

Assuming Gaussian noise, the likelihood of having data $d(t)$ given the presence of a GW signal $h(t)$ is given via the usual GW likelihood (4.18).

17.1.1 Joint Parameter Estimation

When the noise component in Eq. (17.1) is made of multiple signals, the likelihood described by Eq. (4.18) becomes

$$p(d|\boldsymbol{\theta}, \mathcal{H}_{\text{GW}}) \propto \exp \left[-\frac{1}{2} \left\langle d - \sum_{i=1}^N h(\boldsymbol{\theta}_i) \middle| d - \sum_{i=1}^N h(\boldsymbol{\theta}_i) \right\rangle \right], \quad (17.2)$$

where we just expanded the expression for $h(t)$ compared to the previous expression. Here, N is the total number of signals, and $\boldsymbol{\theta}_i$ represents the set of parameters describing the i^{th} GW signal so that $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_i, \dots, \boldsymbol{\theta}_N\}$ in this case.

Using Eq. (17.2) gives rise to the method of JPE, where we jointly look for $\{\boldsymbol{\theta}_i\}_{i=1,\dots,N}$. Since each $\boldsymbol{\theta}_i$ is a set of 15 parameters for spinning BBHs, a set of 16 parameters for NSBHs, and a set of 17 parameters for BNSs⁴, it means that, if we have X BBHs, Y NSBHs, and Z BNSs (so $N = X + Y + Z$), the parameter space has $15X + 16Y + 17Z$ parameters to explore. In the end, this means that as soon as we consider two signals, the parameter space grows to at least 30 dimensions, which is already challenging with our traditional methods, showing the difficulty of analyzing several signals jointly.

In addition, we can assume there is some uncertainty on the total number of signals N in the data. In this case, it is also possible to sample over N and the signal types, effectively allowing for any number of signals to be present in a given data stretch.

The problem of joint analysis of several signals has already been looked at in other contexts, such as the characterization of the nearly monochromatic signals coming from white dwarfs in LISA [419], BNSs in the BBO [374], or supermassive BHs in PTA searches [420]. However, the different signal characteristics looked for in these various context makes the methods different from one case to the other.

In this work, we will only consider the possibility of having two signals in the data. So, we write the data as

$$d(A, Bt) = h_A(t) + h_B(t) + n(t), \quad (17.3)$$

where we just denote the signals by A and B , without any importance on which signal is A and which signal is B .

In this case, the likelihood (Eq. (17.2)) takes the particular form

$$p(d(A, B) | \boldsymbol{\theta}, \mathcal{H}_{\text{GW}}) \propto \exp \left[-\frac{1}{2} \left\langle d(A, B) - h_A(\boldsymbol{\theta}_A) - h_B(\boldsymbol{\theta}_B) \middle| d(A, B) - h_A(\boldsymbol{\theta}_A) - h_B(\boldsymbol{\theta}_B) \right\rangle \right]. \quad (17.4)$$

In principle, if the sources have the same nature, the labels A and B are interchangeable during the sampling, making the likelihood symmetric in two events. In our algorithm, we do not impose any conditions on the parameters to break this symmetry. As a consequence, it needs to be done in a post-processing step, as we need to assign drawn samples to the correct event. Sometimes, not accounting for this condition leads to bimodalities. In this work, we use a time ordering condition, taking the samples for event A to be those that arrive first in time and the samples for event B to be those arriving second in time. In future work, this condition could directly be imposed in the algorithm by having a conditional prior such that the time of arrival of one event is

⁴Typically, a BBH is described by two mass parameters, 6 spin parameters, a distance parameter, the inclination, 2 parameters for the sky location, the merger time, the phase of coalescence, and the polarization angle. Usually, for each NS present in the system, one also adds tidal deformability. However, for BNSs, the dimensionality could increase even further if higher-order tidal contributions, spin-induced quadrupole effects, or resonant effects are accounted for. In addition, formally, when parameter estimation is performed, one can also add the noise-related calibration parameters, which would further increase the dimensionality of the parameter space.

always before the other. We also note that the condition could be imposed on different parameters, such as the chirp mass for example.

17.1.2 Hierarchical Subtraction

In HS, the idea is not to fit the two signals at once but instead combine individual signal analyses and subtraction of best-fit parameters. Therefore, we start by running a single event parameter estimation analysis on the data $d(A, B)$ to get the characteristics of the dominant signal. If we label by A the loudest signal, we can denote the best-fit parameters (typically the maximum likelihood parameters) $\hat{\theta}_A$, and the waveform corresponding to this signal

$$\hat{h}_A(t) = h(t, \hat{\theta}_A). \quad (17.5)$$

Using this, we can get the data for signal B given signal A by subtracting the best-fit template

$$d(B, r_A, t) = d(A, B, t) - \hat{h}_A(t), \quad (17.6)$$

where r_A are the residuals of signal A due to the imperfect subtraction. This is an approximate data strain for the second event in the data since the maximum likelihood parameters used to model the first event are prone to errors, with errors coming from the modeling itself but also from neglecting the presence of the second event when characterizing the first one.

We can then analyze $d(B, r_A, t)$ to get the parameters for event B , leading to a posterior distribution for the two events. In principle, if the bias on the first recovery is not too important, then the posteriors on the second event should also be correct. However, this approach is less robust than the JPE, where we correctly account for the presence of several events.

In Ref. [362], the authors also suggest a way to correct the bias due to the individual characterization of two signals⁵. Once the two signals are analyzed separately, we can use the two best-fitting posteriors to evaluate the bias in the model reconstruction for each signal. The estimated biases can then be applied as a correction factor to the best-fitting parameters. We could then redo the subtraction of each event and analyze it again, but now with a subtracted signal closer to the real one, reducing the possible bias in the recovered posterior. Though this method is attractive, it requires multiple parameter estimation runs, which are expensive in a 3G detector context. The computation of the biases also requires solving a combination of Fisher matrices and numerical derivatives, making it a non-trivial operation.

⁵In their paper, they also account for the possible confusion background due to the sum of all the mergers going on in the background.

Parameter	Population generation	Prior
Primary component mass	PowerLaw + Peak [40]	/
Mass ratio	PowerLaw + Peak [40]	$\mathcal{U}(0.1, 1)$
Chirp mass	/	$\mathcal{U}(4 M_{\odot}, 200 M_{\odot})$
Redshift	Oguri's fit [389] + rescaling	/
Luminosity distance	/	Uniform comoving volume [1, 100] Gpc
Spin amplitude 1	$\mathcal{U}(0, 1)$	$\mathcal{U}(0, 1)$
Spin amplitude 2	$\mathcal{U}(0, 1)$	$\mathcal{U}(0, 1)$
Tilt angle 1	Uniform in sine	Uniform in sine
Tilt angle 2	Uniform in sine	Uniform in sine
Spin vector azimuthal angle	$\mathcal{U}(0, 2\pi)$	$\mathcal{U}(0, 2\pi)$
Precession angle about angular momentum	$\mathcal{U}(0, 2\pi)$	$\mathcal{U}(0, 2\pi)$
Inclination angle	Uniform in sine	Uniform in sine
Wave polarization	$\mathcal{U}(0, \pi)$	$\mathcal{U}(0, \pi)$
Phase of coalescence	$\mathcal{U}(0, 2\pi)$	$\mathcal{U}(0, 2\pi)$
Right ascension	$\mathcal{U}(0, 2\pi)$	$\mathcal{U}(0, 2\pi)$
Declination	Uniform in cosine	Uniform in cosine
Time of coalescence	Uniform over a year (second precision)	$\mathcal{U}(t_{inj} - 0.1, t_{inj} + 0.1)$

Table 17.1: Overview of the functions used to generate the different parameters for the BBH population and the priors used for the parameter estimation recovery.

17.2 Setup of the Analyses

Due to the computational resources required to analyze 3G signals, we focus on overlapping BBHs with masses in $[30, 60] M_{\odot}$ ⁶. We use a network of detectors made of one triangular ET with 10 km arm-lengths and a CE detector located at the LIGO-Hanford position and with 40 km arm-lengths. We generate stationary Gaussian noise from the detectors' PSDs, where, for ET, we use the ET-D PSD [30, 31], and for CE, we use the projected PSD from Refs. [50, 427]. We then inject two simulated BBH signals into the artificial noise. For this study, we take a lower cutoff on the signals of 20 Hz. A representation of the waveforms obtained by the addition of two BBHs can be found in Fig. 17.1. One sees that the final signal has a non-trivial shape, illustrating the risk of biases when not accounting for the presence of two signals. In addition, one also sees that depending on the relative SNR of the signals, the observed bias is different. For the case where one signal has a significantly higher SNR than the other ($\sim \times 2$, top panel), the quieter signal will somewhat bias the signal but the observed waveform will

⁶The higher masses are chosen to not have a signal with a too long durations while still enabling overlap for the difference in arrival times used in this work.

resemble mostly the loudest signal. On the other hand, for signals with close SNRs (bottom panel), we see that the deformation of the signal can be more complicated without really having a dominant signal (except for the fraction of a second where one signal has merged and the other is still merging). Based on these observations, one can expect HS to be more effective when one signal clearly dominates over the other.

In our study, we simulate 55 such mergers. To produce these high-mass signals, we sample the component masses from the Power-law + Peak distribution from [40] but keep only the systems fulfilling the mass requirement. In addition, the events are sampled in redshift according to the merger rate density reconstructed from Oguri's fit [389]. The sky location is drawn to be uniform on the sky, and the spin parameters are picked from an isotropic distribution. For overlapping signal events, the coalescence time of the first event is drawn from a uniform distribution spanning over an entire year, while the second event is placed 0.1 seconds later. An overview of the functions used to make the binaries and the priors used for the analyses is given in Table 17.1.

Since the SNR of the signals can reach hundreds to thousands in an ET and CE network, and such high values make the computation time even longer, we decided to rescale the SNR to take values constrained between 8 and 50. This is done by adjusting the luminosity distance. However, since we expect the SNR ratio between the events to play a role, we try to keep this ratio as close as possible to the original one. So, if the loudest signal has a value above 50, we rescale it to take one between 45 and 50 (this value is drawn randomly from a uniform distribution). Then, we rescale the quieter signal with the same factor. If this value is below 8, then we choose a new scaling factor to bring the SNR back between 8 and 13 (once more using a uniform distribution). Each system is then analyzed once without additional noise and once injected in Gaussian noise generated from the PSDs.

For the different runs, we choose fixed priors for the various parameters. The right column of Table 17.1 gives an overview of the priors used for the different parameters. In particular, we take a uniform prior on chirp mass (\mathcal{M}_c , Eq. (2.50)) and mass ratio ($q = \frac{m_2}{m_1}$), with bounds of $[4, 200] M_\odot$ and $[0.1, 1]$. We also take a uniform in comoving volume prior for the luminosity distances, with bounds going from 1 to 100 Gpc. These priors are adapted to cover any possible signal present in our set of data. The priors for the other parameters correspond to the usual priors taken for BBHs. When doing the JPE and HS runs, the priors are the same for the two events, and no conditions related to the signals (for example, time ordering of the signals or enforcing a heavier one) are added.

We note here that an alternative approach is to use narrower priors informed by the results of low-latency searches [380, 386], which very likely could only be applied for a couple of parameters such as the chirp mass and the coalescence time. In [380], the authors show that matched filtering pipelines and unmodelled searches can pick up overlapping signals with reasonable accuracy. They also suggest some enhancements to make the pipelines even more suited for the challenge of overlapping signals detection. In addition, they show that for most of the overlapping signals, the error on the chirp mass

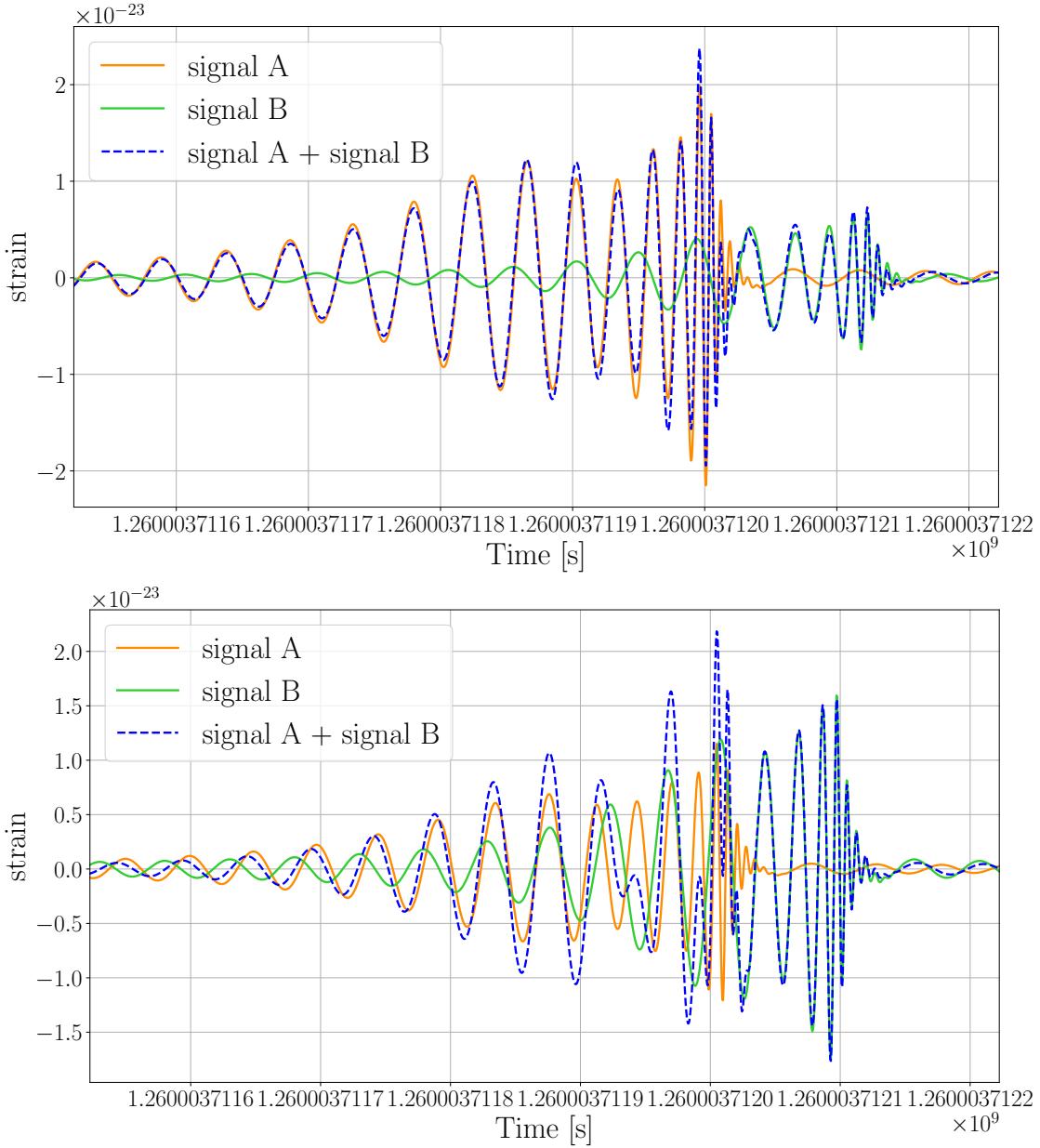


Figure 17.1: *Top:* Representation of two overlapping BBHs and the underlying signals for signals with different SNRs ($\text{SNR}_A = 46.1$ and $\text{SNR}_B = 22.2$). In this case, the sum of the two signals is mostly dominated by the loudest signal, while the effect of the quieter signal takes up only after the loudest signal has merged. *Bottom:* Representation of two overlapping BBHs and the underlying signals for signals with similar SNRs ($\text{SNR}_A = 34.3$ and $\text{SNR}_B = 30.2$). In this scenario, no signal dominates over the other for the entire event duration, and the overlapping signals have more complicated features.

is not much larger than for the non-overlapping case, even if for the occurrences where the signals are very close in time, the error increases. Therefore, using such searches to set narrower priors is a realistic alternative. However, currently, they also seem to contain risks, as an increased difference in the value recovered for some parameters can happen, and taking too narrow a prior could lead to the exclusion of the actual value from it. In the end, additional developments are needed to ensure that using these results to narrow down the initial priors is viable.

To have a basis of comparison, we also perform the parameter estimation of the individual signals. This is done by using the same priors as the one explained above but injecting only one of the two signals in the noise. The single parameter estimation runs are denoted SPE.

All the parameter estimation runs are performed using BILBY [149] with the DYNESTY [151] sampler. For the JPE runs, we add our own adapted joint likelihood in the package to keep a consistent framework.

17.3 Results and Discussions

In this section, we show the results of the different approaches. We first compare the HS approach with SPE. Then, we compare JPE with SPE and HS. Here, we focus on the results of the analyses performed with noise. The conclusion in the no-noise case is similar and can be found in Sec 17.4. For all the figures presented in this work, when plotting individual event results, we represent by a dot and label as “loud” events those that are the loudest in the pair, and by a triangle and label as “quiet” the quieter ones.

17.3.1 Hierarchical Subtraction

We start with discussing the result of the analysis for the HS approach. In almost all HS runs, the first PE stage picked up the signal with higher SNR. There were two cases where it instead picked up the quieter signal; these were instances where the SNRs of the injected signals were close to each other. In such a case, the signal picked first is not the same with and without noise. A representative example of the posteriors can be found in Fig. 17.2. While the widths of the distributions match the ones obtained for SPE runs closely, in most cases they show bias in the recovered parameters.

When comparing the HS and SPE recoveries of the parameters, one sees that the HS recovery is nearly always biased. However, this bias seems more pronounced when the injection is done in noise. For example, in Fig. 17.2, one sees the zero-noise HS recovery is close to the one for SPE with noise, while the recovery with noise is off. This shows that the mismodelling of the noise (due to an additional event) is present, as

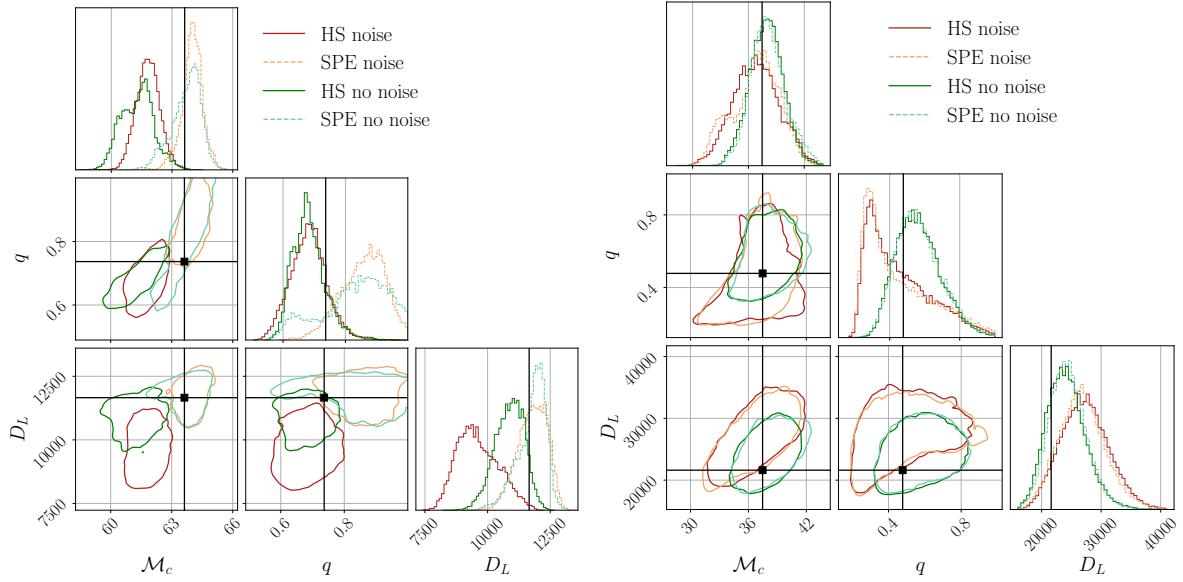


Figure 17.2: Comparison of the posteriors for the HS and the SPE methods with and without noise. We represent the chirp mass (\mathcal{M}_c), the mass ratio (q), and the luminosity distance (D_L). *Left:* Case where HS is worse than SPE. HS posteriors are shifted significantly compared to the injected value and SPE posterior recovery. *Right:* Case where HS is close to SPE, with equivalent posteriors recovered in the two cases.

one could expect from previous works on biases in overlapping signals [319, 360–362]. However, this does not necessarily mean that all parameters are off and the first signal’s characteristics can be recovered.

In Fig. 17.3, we can see mismatches between the injected and recovered waveforms, comparing HS and SPE cases with and without noise. The mismatch is defined as $1 - \bar{M}$, where \bar{M} is the match between the waveforms, defined in Eq. (4.13). The mismatch represents the dissimilarity between two waveforms. High values mean a major disagreement between the two waveforms, and smaller values mean that the waveforms agree well. In Fig. 17.3, we see that the average mismatch throughout the detections is always low, below 0.02. As expected, the presence of another signal leads to worse waveform recovery for HS compared to SPE (most points are below the diagonal). In our data set, the worst recovery of the signals occurs for overlapping signals with similar SNRs. Note that the zero-noise case (see Sec 17.4) shows a clearer difference between SPE and HS recoveries. It is expected to have a larger difference in this case because the effect of the unmodelled signal is stronger when there is no noise since it is the only source of uncertainty in the signal.

By looking at the difference between the median of the recovered posterior and the injected value, normalized by the injected value, we can quantify the offset in the recovery ($\Delta_{\mathcal{M}_c} = \frac{|\mathcal{M}_c^{\text{inj}} - \mathcal{M}_c^{\text{rec}}|}{\mathcal{M}_c^{\text{inj}}}$, where “rec” stands for recovered and “inj” stands for injected.). This is represented Fig. 17.4 for the chirp mass recovery with noise and is representative of

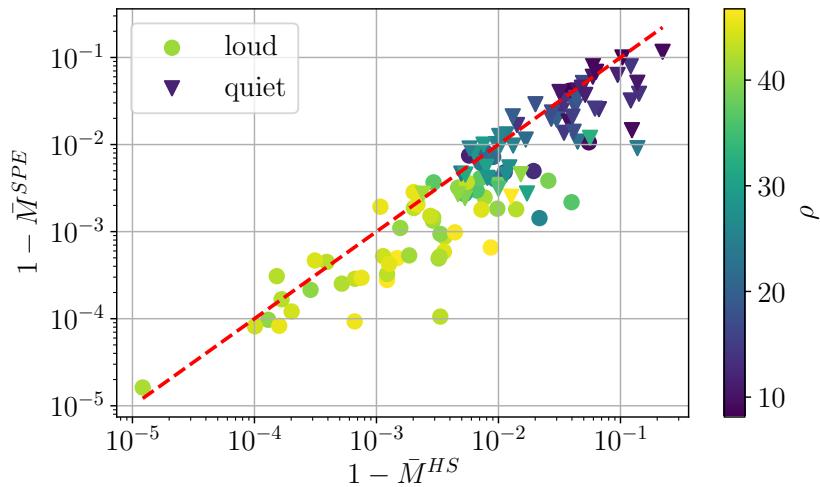


Figure 17.3: Mismatch between the injected and recovered (maximum likelihood) waveforms for the SPE analysis versus the HS analysis, with noise. The red dashed line represents the diagonal where the mismatch is the same. HS is worse at signal recovery, as most points fall below the diagonal. This is expected since HS does not model the noise properly since it neglects the presence of one of the signals.

all the parameters. HS shows higher offsets in 71% of the louder events and 51% of the quieter ones. As expected, the offset is larger for HS compared to SPE. For the first signal recovered, since we have an unmodeled signal in the detectors, the noise properties are not modeled properly. Therefore, generally, a larger offset is observed for the louder signal compared to the secondary one. However, when strong deviations are present for the first signal, they can reverberate in the second, also leading to worse recoveries for this event.

Finally, Fig. 17.5 shows how the widths of the 90% confidence intervals of the posteriors for HS compare to SPE, normalized by the injected value ($\delta_{\mathcal{M}_c} = \frac{\sigma_{\mathcal{M}_c}}{\mathcal{M}_c^{\text{inj}}}$, where $\sigma_{\mathcal{M}_c}$ represent the width of the 90% confidence interval). We observe that the widths of the distributions are consistent between the two, even though the recovery is biased.

It is interesting to compare how doing successive parameter estimation steps affects the results. After reconstructing the quieter of the two signals, we subtract it from the initial data and do parameter estimation again. In principle, it should result in a better recovery of the louder signal than the original parameter estimation run.

The top panel of Fig. 17.6 shows the mismatch for the recovered dominant waveform after the first and third HS runs. The match after the third run is better in 62% of the cases, compared with 50% expected if the procedure had no effect at all. This small effect is also observed on the offset plot – see Fig. 17.6, bottom panel – where the third HS step leads to better results in the same proportions. Even if the results get better for some events, it also leads to worse recoveries for other cases, and only a few of the other events have comparable results between the first and third HS steps. Therefore, it

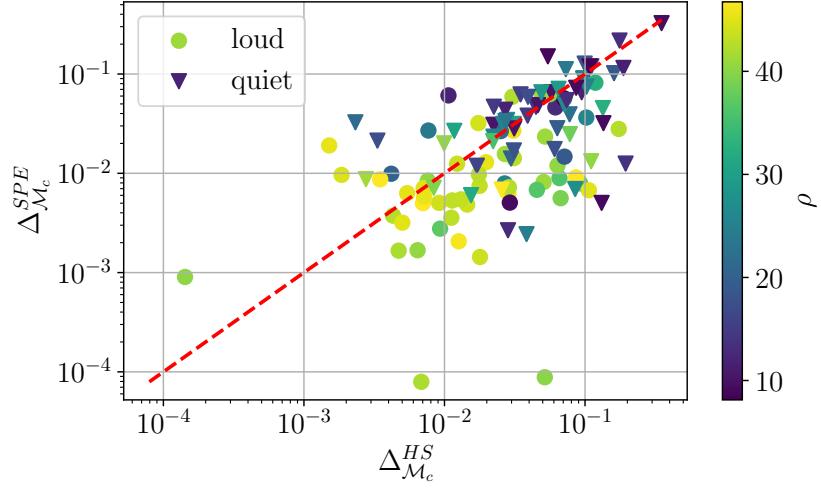


Figure 17.4: Representation of the offset of the recovered parameters, with noise for HS and SPE. Plotted is the difference between injected value and the median of the recovered value, normalized by the injected value. HS tends to give a higher deviation than SPE.

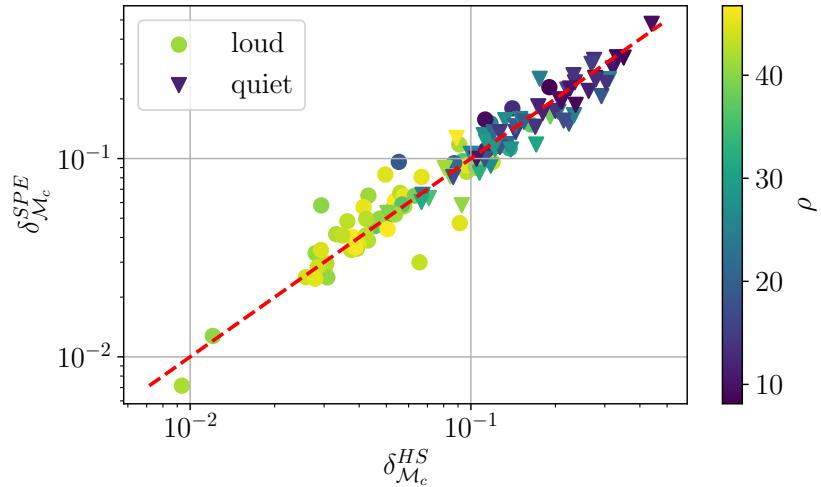


Figure 17.5: Comparison of the normalized width of the 90% confidence interval for the chirp mass for the HS recovery and the SPE recovery. We see that the width of the recovered distribution is largely unaffected by the presence of another signal.

does not seem like simply applying successive HS steps converges to unbiased posteriors. More sophisticated approaches appear to be needed, like an estimation of the offset as suggested in [362] and briefly explained in Sec. 17.1.2.

In the end, our HS runs confirm previous observations [319, 360, 361]: doing parameter estimation for one signal neglecting the other can lead to significant biases when the two signals merge very close to each other. In addition, we have also shown that once the first signal is subtracted, analyzing the second one with parameter estimation is less prone to deviations, even if the subtraction of the first event is not perfect.

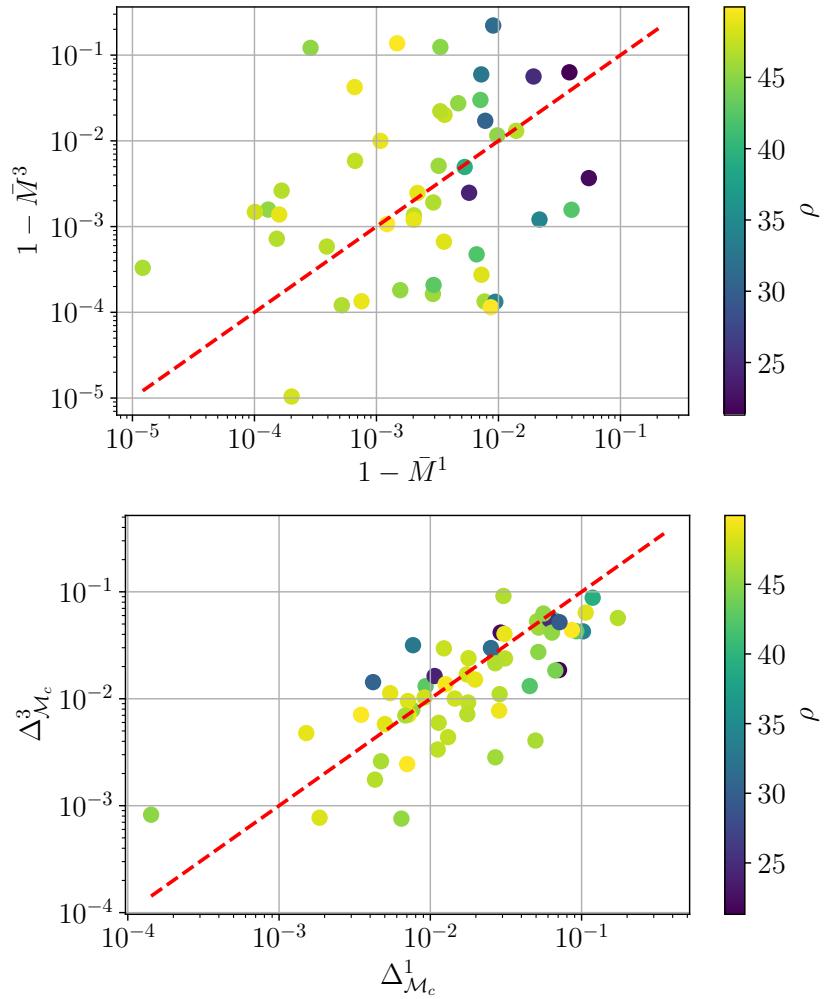


Figure 17.6: *Top:* Comparison of the mismatch between the injected and recovered waveforms for the first (horizontal axis) and third (vertical axis) steps of HS. The mismatch after the third step is lowered in 62% of the cases. *Bottom:* Comparison of the offset in recovered chirp mass between the first (horizontal axis) and the third (vertical axis) HS steps. The recovered bias after the third step is lowered in 62% of the cases.

17.3.2 Joint Parameter Estimation

In what follows, we discuss the results obtained from the analysis of the overlapping signals using JPE. We focus on the results obtained after the time ordering of the samples (described in Sec. 17.1) is done, as these are the samples where effectively one set of posteriors is matched with one signal and the other set with the other signal.

For the JPE recovery, more diverse scenarios are possible. We show three main cases in Fig. 17.7: one where the recovery is equivalent to the one from SPE, one where JPE has smaller bounds on the posteriors for one or more parameters, and one where JPE has trouble fitting the signal properly and offsets can occur. While more in-depth studies are required to fully comprehend this behavior, it seems like it could be originating from the mixed term of the two signals present in the likelihood modified to account for multiple signals. In some cases, the narrower posteriors could be offset compared to the injected value, pushing it out of the 90% confidence interval.

For all the events, we compare the mismatch between the maximum likelihood waveforms of the event in the JPE scenario and the SPE and HS cases. This is represented in Fig. 17.8 for the noise cases. Independent of the presence of noise, we find that the mismatch is smaller for JPE than for HS but larger than for SPE. This is what one would expect since JPE accounts for the presence of the two signals and so should lead to smaller deviations. However, fitting two GW signals simultaneously is more complex than analyzing a single signal. Therefore, the SPE measures remain a better representation of the injected signal.

As before, we present the normalized distance between injection and recovery in Fig. 17.9 (for the noise case). When comparing the offset for the JPE case against the SPE case, we find that 45% of the events have a larger offset for JPE than for SPE. On the other hand, when we compare with HS, we find 65% of the events with a lower offset for JPE. This confirms that JPE is better than HS for finding the injected signals (when the two are at 0.1 s of each other in the data). This is indeed what one would expect, as JPE takes care of the mismodeling of the noise but leads to an increased complexity during the analysis.

For the spread in recovered posteriors, contrary to what one had for the HS approach, the normalized width of the 90% confidence interval does not align on the diagonal. Indeed, since we have more varying scenarios, with larger or tighter posteriors in some cases, the spread can be significantly different between the JPE and SPE scenarios. In addition, since there is no significant difference for this quantity between HS and SPE, the relation between JPE and HS is the same as between JPE and SPE. The increased discrepancy between the two approaches is represented in Fig. 17.10. Nevertheless, the posteriors are distributed evenly above and below the diagonal representation, showing that, on average, the posteriors' width is not significantly bigger in one method or the other.

Based on our JPE results, we see that, using an adapted likelihood (17.4), we obtain

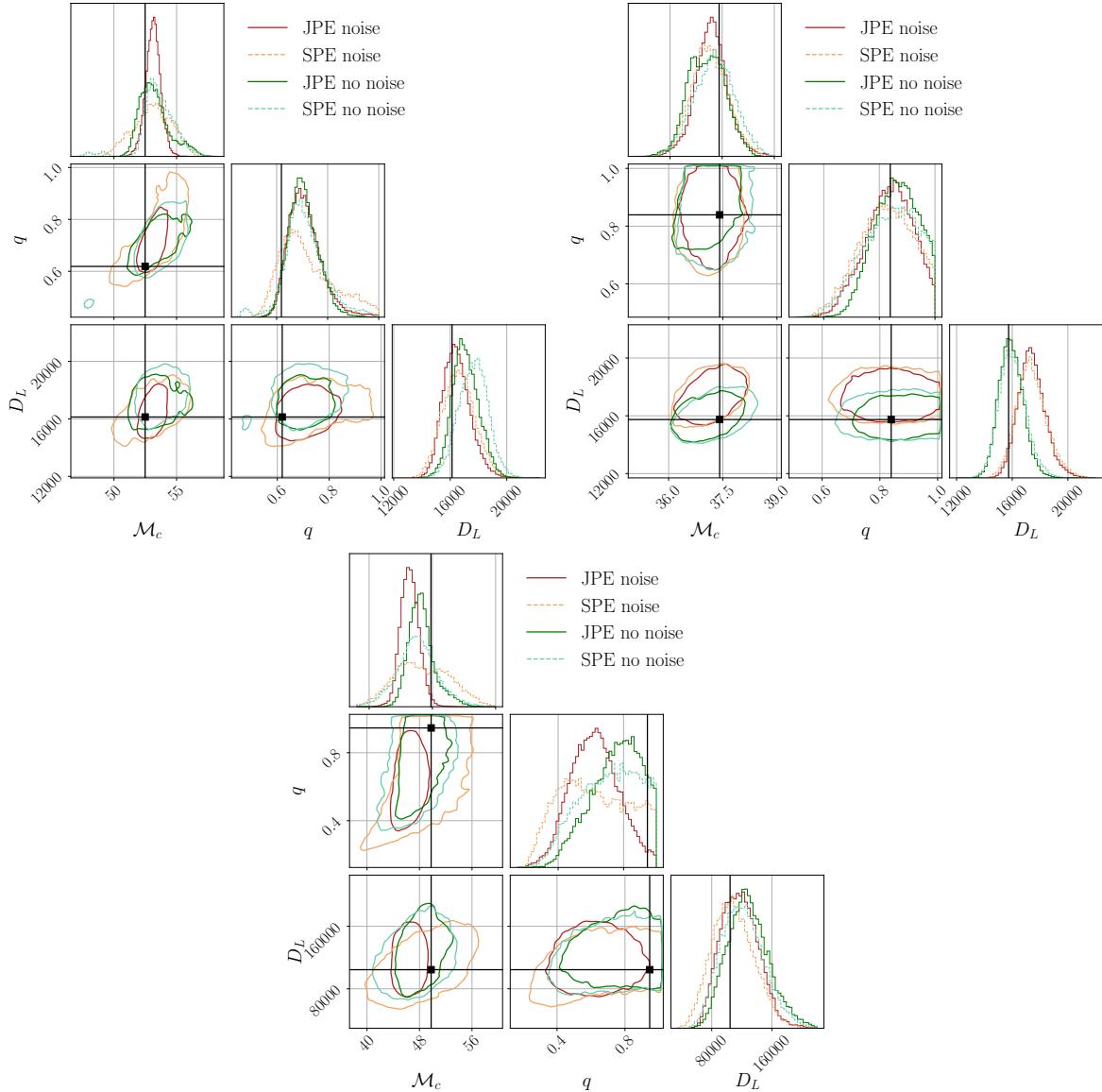


Figure 17.7: Comparison of the recovered posteriors for JPE and SPE with and without noise for different types of recovery. *Top left:* the posteriors recovered with JPE for the noise case is narrower for the chirp mass and the mass ratio compared to the SPE case, while the posterior for the luminosity distance is narrower for JPE in the zero-noise case. *Top right:* JPE and SPE are very close to each other, with equivalent recovery with and without noise. *Bottom:* representation of a case where the JPE recovery is worse than for SPE. We get narrower posteriors, but the peak is shifted out of the 90% confidence region.

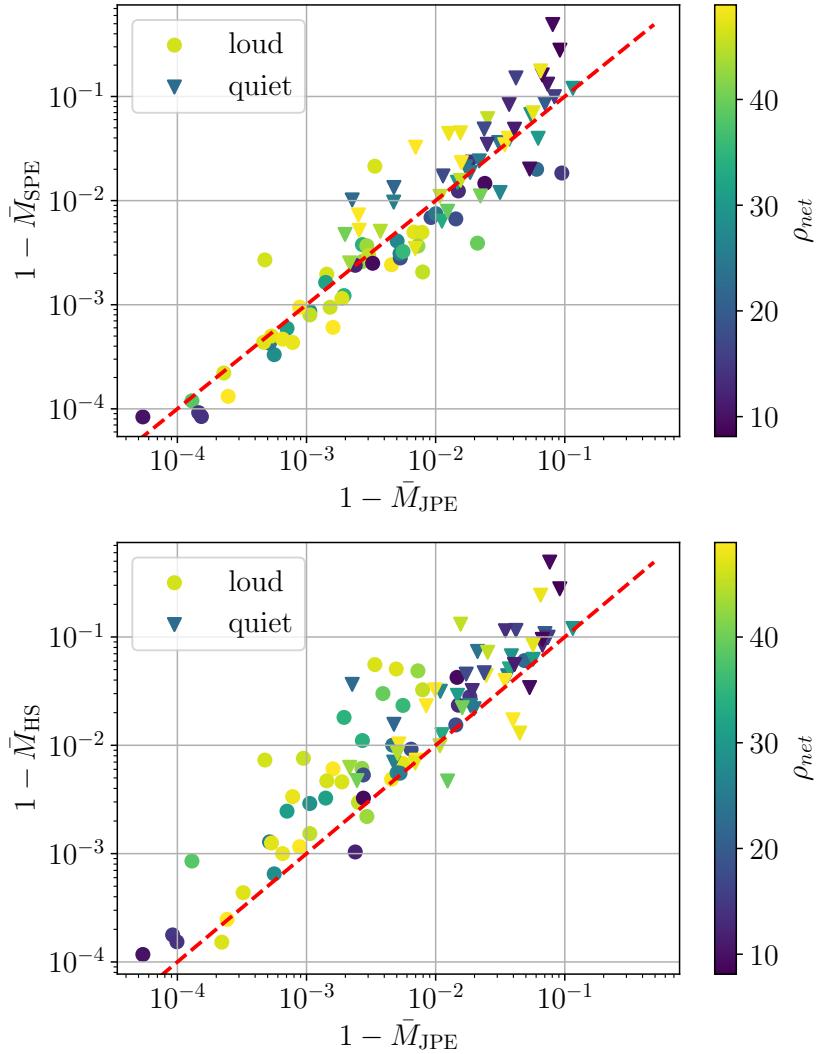


Figure 17.8: *Top:* Comparison of the mismatch between the injected and recovered waveforms for the JPE and the SPE cases with noise. *Bottom:* Comparison of the mismatch between the injected and recovered waveforms for JPE and HS with noise. Overall, the mismatch is higher for JPE than for SPE, while it is lower than for the HS case. This is expected since JPE accounts for the two events in the data, which is better than neglecting one but more complicated than having only one signal present in the data and fitting that signal.

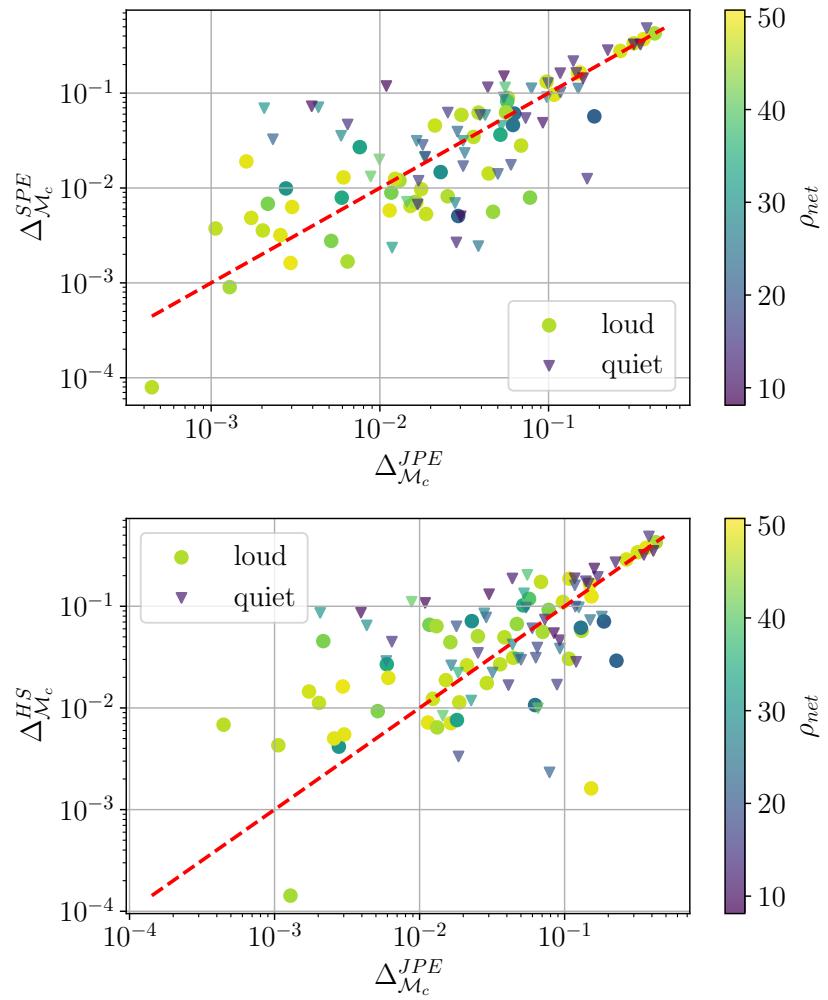


Figure 17.9: *Top:* Comparison of the offset of the recovered posterior for the chirp mass for the JPE and SPE methods. *Bottom:* Comparison of the offset of the recovered posteriors for the chirp mass for JPE and HS. The two plots indicate that the offset is lower for JPE than for HS, due to the better modeling of the noise, while it is still better in the SPE case, where the noise is well modeled, and the problem at hand has a reduced complexity.

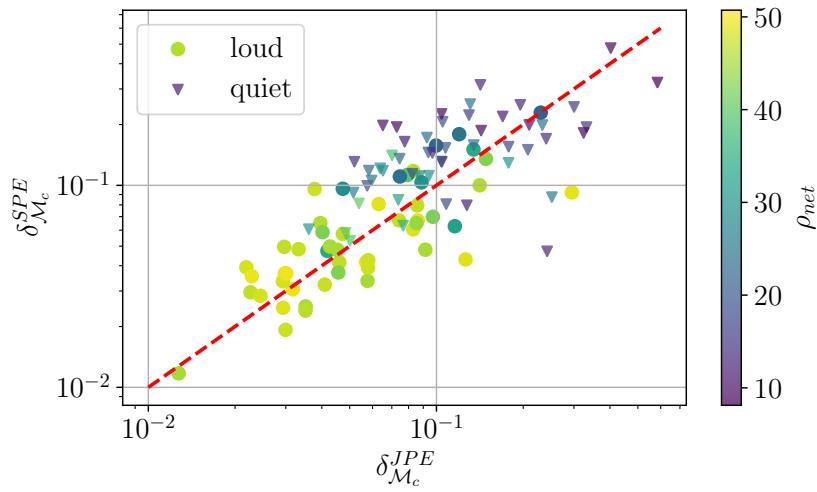


Figure 17.10: Comparison between the normalized width of the 90% confidence interval for JPE and SPE for the noise case. Since the spreads are very close for SPE and HS, the same relation is established for the JPE and HS comparison. Here, we see there are larger differences between the two approaches but, globally, one is not better than the other as we have about 50% of the events above the diagonal and the same proportion below.

better results than by analyzing both signals sequentially when their merger times are very close. However, this comes with the drawback that the computational time and the complexity of the problem are increased, making the approach less stable as can be seen by the wider variations in the widths of the posteriors. On the other hand, in some cases, JPE leads to narrower posteriors compared to SPE. This could be due to the inner product mixing the two signals following the introduction of multiple signals in the likelihood. However, a more in-depth study should be performed on a larger set of events to understand this behavior. This is left for future work.

17.4 Zero-noise Results

In this section, we show results for the posteriors obtained when the injections are analyzed without noise. The conclusions drawn from these experiments are the same as in the noise case, which suggests that our findings are robust. So, they are due to sampling effects and not induced by random noise fluctuations added to the data.

17.4.1 Hierarchical Subtraction

Here, we present the complementary zero-noise result for the HS method.

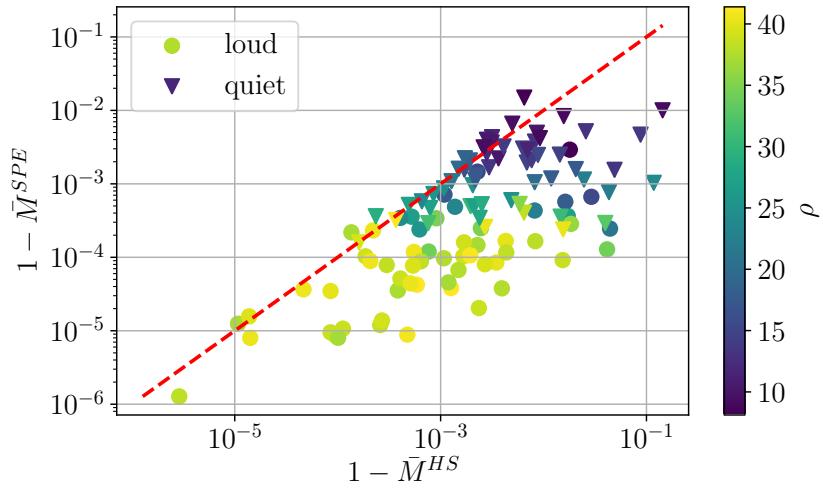


Figure 17.11: Mismatch between the injected and recovered waveforms for the SPE analysis versus HS analysis without noise. The red dashed line represents the diagonal where the mismatch is the same. The difference in waveform recovery between SPE and HS is more pronounced than in the case with noise, and it is clear that the recovery degrades when using the HS approach.

Fig. 17.11 represent the mismatch for HS versus SPE for the zero-noise case. As for the noise case, HS leads to higher mismatches, meaning that the recovered (maximum likelihood) parameters are a worse representation of the injected signals.

Fig. 17.12 represents the offset for the chirp mass for HS versus SPE without noise.

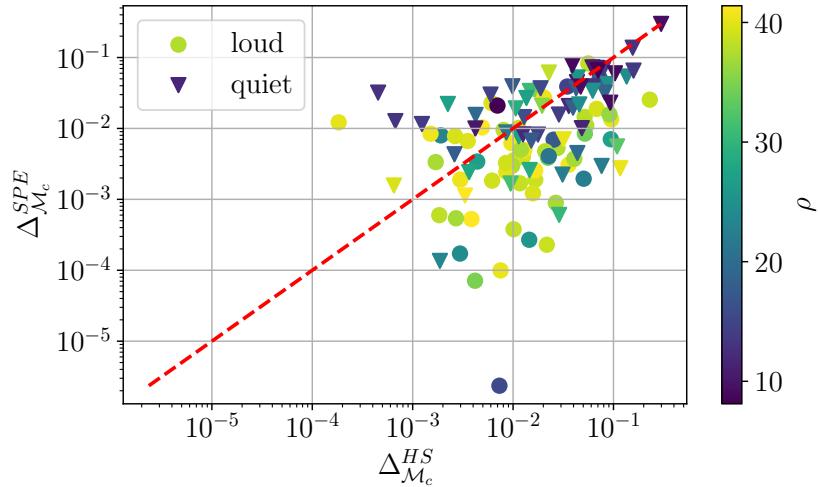


Figure 17.12: Representation of the offset of the recovered chirp mass without noise. Plotted is the difference between injected value and the median of the recovered value, normalized by the injected value for HS and SPE. One sees that offsets are more important for HS.

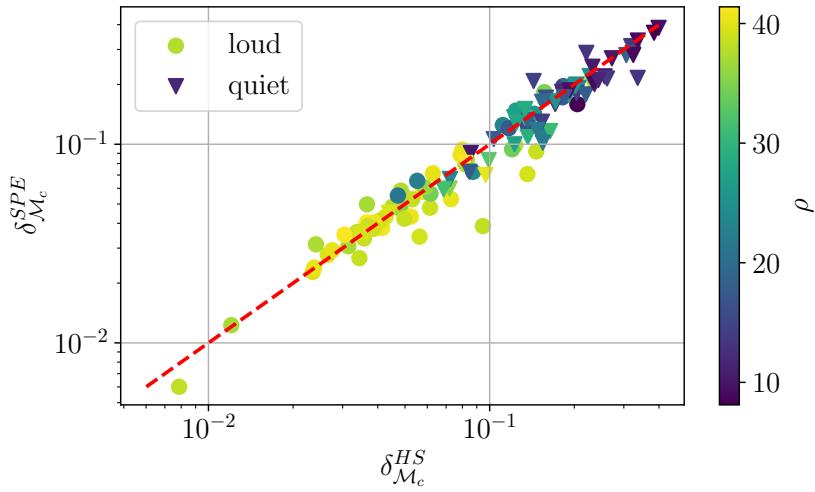


Figure 17.13: Comparison of the normalized width of the 90% confidence interval for the chirp mass for the HS recovery and the SPE recovery, with zero-noise. We see that the width of the recovered distribution is largely unaffected by the presence of another signal.

Here HS shows higher offsets for 74% of the louder events and 57% of the quieter ones. There is no significant difference compared with the injections into noise, with HS still being more prone to deviations in the recovered posteriors.

Fig. 17.13 represents the normalized width of the posteriors for HS versus SPE. Similarly to the analysis with noise, the widths of the distributions are very close to each other.

17.4.2 Joint Parameter Estimation

Here, we present the complementary zero-noise results for the JPE method.

Fig. 17.14 represents the mismatch between the JPE and SPE methods (top) and the JPE and the HS methods (bottom). This also shows the JPE method leads to a better representation of the data than the HS method. However, the increased complexity of the problem leads to a decrease in the accuracy of the recovery. We keep low mismatch values for JPE and SPE.

Fig. 17.15 represents a comparison between the offsets for JPE and SPE, and JPE and HS, for the zero-noise case. Here, one also has a larger offset for the JPE than for SPE (39% of the events have a smaller offset for JPE), and a larger bias for HS as for JPE (57% of the events have a smaller offset for JPE).

Fig. 17.16 represents the normalized width of the posteriors for JPE versus SPE (which is comparable to JPE versus HS since HS versus SPE has widths aligning along the diagonal). There is more variance in this plot than for SPE versus HS. This is because the JPE method is significantly different from SPE, and we have a larger variety of

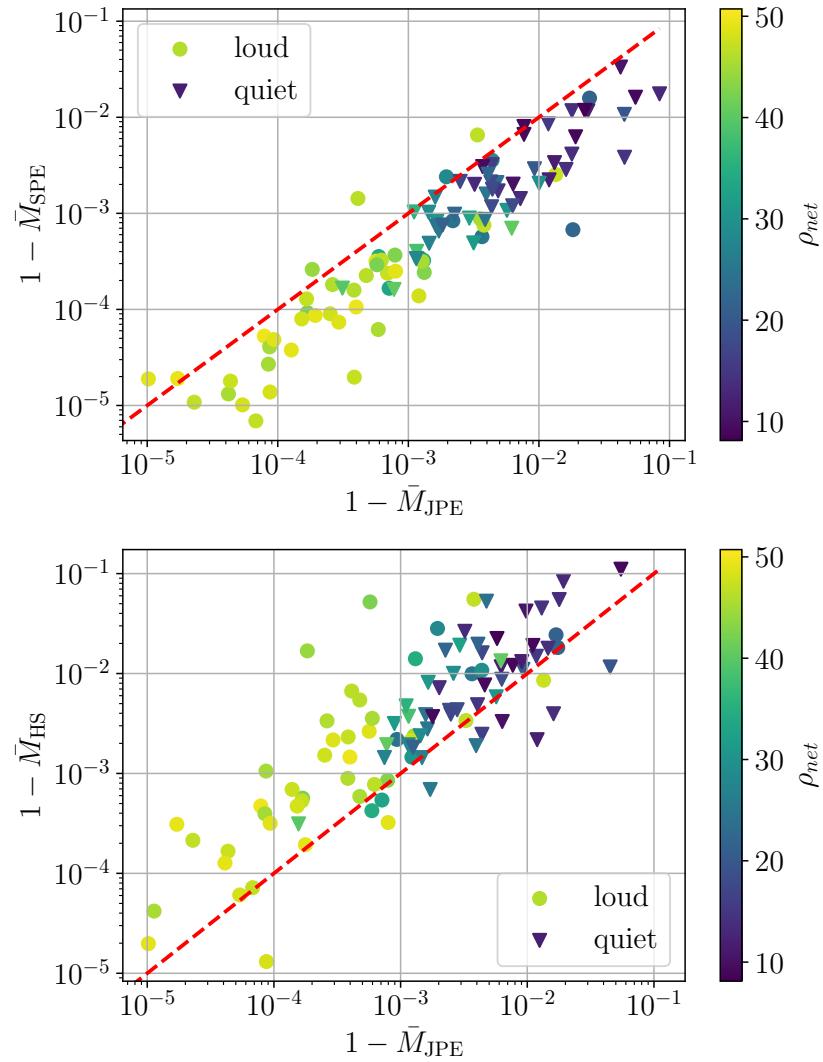


Figure 17.14: *Top:* Comparison of the mismatch between the injected and recovered waveforms for the two events for JPE and SPE without noise. *Bottom:* Comparison of the mismatch between the injected and recovered waveforms for JPE and HS without noise. Overall, it is higher for JPE than for SPE, while it is lower than for the HS approach. This is expected since JPE accounts for the two events in the data, which is better than neglecting one but more complicated than having only one signal present in the data and fitting it.

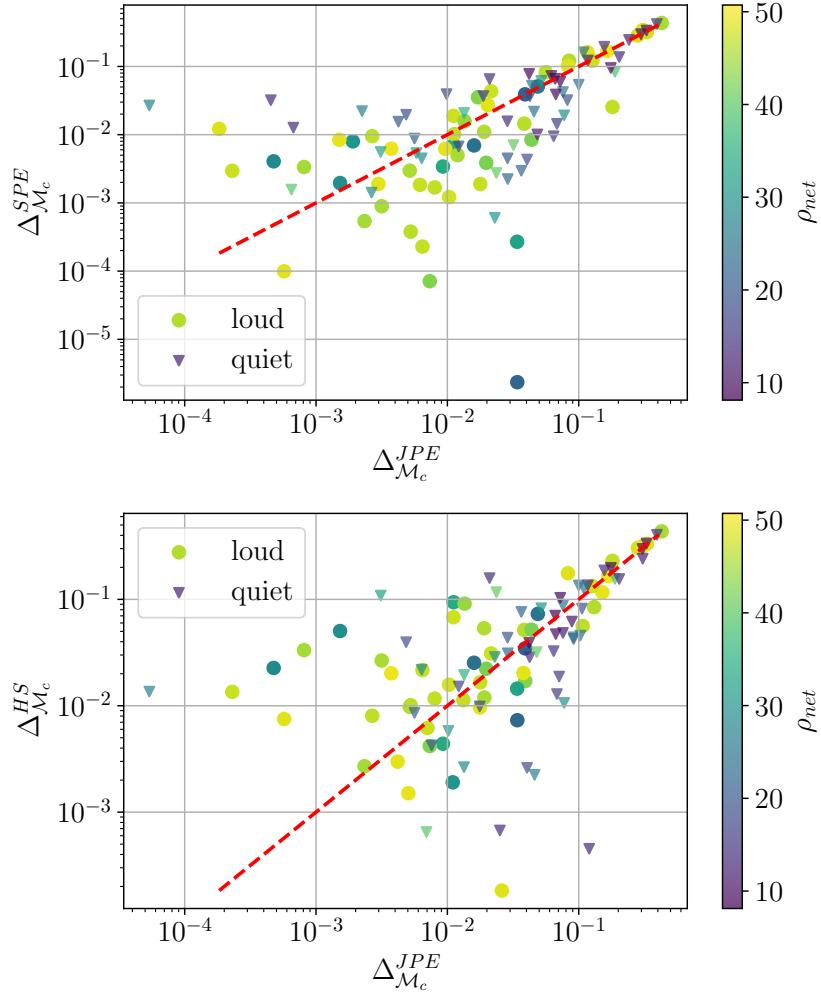


Figure 17.15: *Top:* Comparison of the offset of the recovered posterior for the chirp mass for the JPE and SPE methods in the zero-noise case. *Bottom:* Comparison of the offset of the recovered posteriors for the chirp mass for JPE and HS in the zero-noise case. The two plots indicate that the offset is lower for JPE than for HS, due to the better modeling of the noise, while it is still better in the SPE case, where the noise is well modeled, and the problem at hand has a reduced complexity.

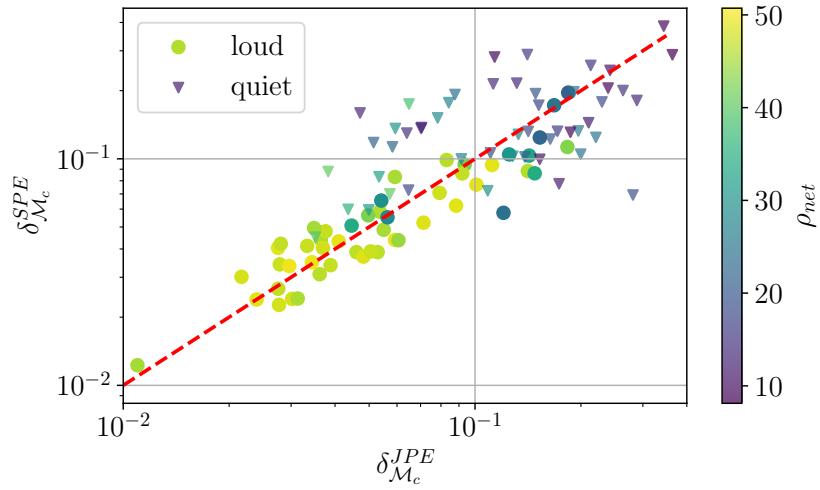


Figure 17.16: Comparison between the normalized width of the 90% confidence interval for JPE and SPE without noise. Since the spreads are very close for SPE and HS, the same relation is established for the JPE and HS comparison. There are larger differences between the two approaches, but, globally, one is not better than the other as the results are more or less evenly distributed around the diagonal.

posteriors. Indeed, sometimes, for JPE, we get broader posteriors but also tighter ones, depending on the characteristic of the two signals in the data. However, the points are evenly distributed on the two sides of the diagonal, showing that, on average, none of the methods has widened posteriors compared to the other.

17.5 Conclusions

In this study, we have presented two ways to perform parameter estimation for two overlapping BBH signals: HS and JPE. We have implemented them and compared them to the usual SPE. Because of limited computational resources, we focus on high mass and medium SNR events; major adaptions to the parameter estimation framework are needed to deal with a larger variety of signals.

For the HS method, we analyze the first signal, then subtract it (more precisely, the maximum likelihood signal) from the data before analyzing the second signal left in the data. On the other hand, JPE analyzes the two signals simultaneously. We note that, since the likelihood is symmetric in the two signals when they have the same nature, we need to add a post-processing step to have samples corresponding to each event. For this purpose, we order the samples in time.

We have applied both methods to a population of BBH mergers to show the feasibility of the two approaches and their respective drawbacks. For HS, as already mentioned

in the literature, we showed that analyzing one signal while neglecting the other can lead to offsets in the recovered posteriors because of the erroneous noise representation. However, even when some deviations occur in the first signal, it does not necessarily impact the recovery of the second signal. We have also shown that there is no significant broadening of the posteriors compared to usual SPE approaches. Therefore, it shows that HS suffers from biases due to overlapping signals for signals measured at very close merger times, which can lead to an incorrect parameter inference.

For JPE, we have shown that the offset in recovered posteriors is smaller than for HS while remaining higher than for SPE. This is understood as JPE having a correct noise representation for the two signals, which HS lacks. On the other hand, solving the likelihood for JPE means we explore a 30-dimensional parameter space, making the analysis more complex and computationally challenging than SPE. However, the recovered width of the posteriors is, on average, the same as for SPE. But while the average is the same, we are confronted with posteriors that can be narrower or broader in JPE compared to SPE. This behavior seems to originate from the cross-term between the two signals entering the likelihood when adapting it to account for multiple signals. More extended studies are needed to understand this effect.

Overall, our results indicate that current techniques for a joint likelihood approach are not yet at their best, and options for making the sampling more efficient could be implemented in future work. For example, one could impose the time ordering (or chirp mass hierarchy) directly during the sampling by imposing that the arrival time of one event is smaller than that of another event. This would prevent the sampler from confusing the two events and enable it to converge more easily. Another possibility could be to use narrower priors motivated by the output of the search pipelines. This is possible for the chirp mass and the arrival time but still contains risk as the search pipelines themselves can provide inaccurate point estimates for some critical parameters [380].

One of the main issues with the methods suggested here is the computational time required, as the data analysis takes up to a few months for lower-mass overlapping BBH mergers. This would make it extremely hard, if not impossible, to keep up with the detection rate of the 3G-detector network. However, methods exist to speed up traditional parameter estimation methods, such as relative binning [354–356] or adaptive frequency resolution [353]. These methods could be adapted to overlapping signals in future work to reduce the computational time, enabling one to analyze other types of systems and to use a lower minimum frequency to get closer to the real 3G scenario. A drastically different approach that could help in the analysis of such signals in the future is ML, where major progress has been made in the parameter inference for single compact binary coalescences [297, 300]. In parallel to this work, a proof-of-concept study applying ML techniques to overlapping signals has shown that it is possible to extract posteriors using normalizing flows with a reasonable precision [428]; see Chapter 18 of this thesis.

To this end, we believe this work makes a first step towards the analysis of overlapping CBC signals, which will be crucial to analyze GW data in the 3G detector era, as

overlaps will become quite common.

18

NORMALIZING FLOWS AS AN AVENUE TO STUDY OVERLAPPING GRAVITATIONAL-WAVE SIGNALS

In Chapter 16 and Ref. [319], it has been established that neglecting the presence of a second signal in the data can lead to biases in the recovered posteriors. This has also been confirmed by other works [360–363]. In Chapter 17 and Ref. [426], we have shown the first steps towards the complete analysis of overlapping signals by comparing two approaches: HS and JPE. The second method is less prone to biases but computationally heavier. Additionally, the method’s current implementation suffers from some instabilities, and upgrades are needed to be entirely reliable. Moreover, the method is computationally heavy, requiring between 20 and 30 days to analyze two high mass BBHs¹. In the 3G era, more than 10^5 CBC mergers are expected [319]. So, having analyses taking several weeks to complete is not an option.

In Ref. [426] and Chapter 17, we suggested several methods to speed up our analyses. Some consist in speeding up the traditional Bayesian methods, using adaptive frequency resolution [353] or relative binning [354–356]. Others imply using conceptually different techniques, such as quantum computing [358] or ML [293–299].

In this chapter, based on Ref. [428], we demonstrate, as a proof-of-concept, how overlapping signals can be analyzed with ML-based techniques. We focus on overlapping BBH signals in a network of 2G detectors. Our ML method relies on normalizing flows [429–431]. We show that ML is a viable approach for analyzing overlapping GW signals, even if further developments are required to obtain optimal performance.

18.1 Machine Learning for Overlapping Gravitational Waves

The use of ML in GW data analysis has been growing over the last years, owing to its speed after training and flexibility; see Ref. [62] for an overview. A subset of these methods falls under the umbrella of simulation-based inference [432] and are developed to perform parameter estimation for CBC signals [293–299]. In particular, Refs. [296–298]

¹This would be the lower bound on the analysis time. Lower mass signals have a longer duration and would therefore need a more extended analysis time.

use normalizing flows to get posterior distributions for BBH parameters. Such methods have been shown to have results close to MCMC and nested sampling. Our approach is somewhat similar to theirs, with some notable differences explained below.

We use *continuous conditional normalizing flows*, a variant of normalizing flows suited for probabilistic modeling and Bayesian inference problems. One can find a detailed overview of normalizing flows in Sec. 6.3. Here, we just give a brief overview to outline the differences between our approach and previous normalizing flow-based methods. Due to the recursive and continuous nature of these models, their memory footprint can be quite small [242]. These qualities allow for extensive training on home-grade GPUs while retaining the ability to capture complex distributions.

Normalizing flows are an ML method through which a NN learns the mapping from a simple *base* distribution $p_u(\mathbf{u})$ (a Gaussian, for example) to a more complex *target* distribution $q(\boldsymbol{\theta})$ (see Sec. 6.3.1 for a detailed introduction). This is done through a series of invertible and differentiable transformations. They are summarized by a function $g(\boldsymbol{\theta})$. However, in our case, it is not sufficient to go from one distribution to the other. We also need to do this conditionally on the GW data we wish to analyze. To account for this, we use *conditional normalizing flows* [238], where the transformation functions are dependent on the data d (hence, $g = g(\boldsymbol{\theta}, d)$; see Sec. 6.3.3). A major difference with [238] is that our base distributions are kept static; experiments on toy models did not show any benefits in having conditional priors. Thus, our model $g(\boldsymbol{\theta}, d)$ is a trainable conditional bijective function that transforms a simple 30-D Gaussian into a 30-D complex distribution. The bijectivity allows us to express and sample $q(\boldsymbol{\theta}|d)$ in terms of $g(\boldsymbol{\theta}, d)$ and $p_u(\mathbf{u})$ via the change of variable theorem:

$$q(\boldsymbol{\theta}|d) = \left| \det(J_{g^{-1}}(\boldsymbol{\theta}, d)) \right| p_u(g^{-1}(\boldsymbol{\theta}, d)), \quad (18.1)$$

where $\det(J_{g^{-1}}(\boldsymbol{\theta}, d))$ is the determinant of the Jacobian $J_{g^{-1}}(\boldsymbol{\theta}, d)$ of the transformation. We train the model by minimizing the forward KL-divergence (see Sec. 6.3.1), which is equivalent to maximum likelihood estimation [431, 433]. As noted by [298], $q(\boldsymbol{\theta}|d)$ should cover the actual (Bayesian) posterior $p(\boldsymbol{\theta}|d)$, and asymptotically approach it as training progresses due to the mode-covering nature of the forward KL-divergence.

A distinctive choice of our method is the continuous nature of the flow, linked to the transformation function itself. A more in-depth explanation of continuous normalizing flows is given in Sec. 6.3.2. Neural ODEs [242] are the foundation of continuous normalizing flows. They are not represented by a stack of discrete layers but a hypernetwork [434]. Hypernetworks can be understood as regular networks where “external” inputs, such as a (continuous) time or depth variable, smoothly change the network’s output for identical inputs. They can thus represent multiple networks or transformations. In [242], hypernetworks are used to represent ODEs and trained by using ODE solvers and clever use of the adjoint sensitivity method. A continuous normalizing flow uses neural ODEs as its transformations.

We will now explain the training of a continuous flow. For clarity, we will use f to

refer to a continuous transformation and g for a discrete one. If $\boldsymbol{\theta}(t)$ represents the samples from the distribution at a given time t , the continuous normalizing flow obeys an ODE when going from t_1 to t_2

$$\frac{d\boldsymbol{\theta}(t)}{dt} = f(t, \boldsymbol{\theta}(t)). \quad (18.2)$$

The change in likelihood associated with this “step” differs slightly from Eq. (18.1) due to the continuous nature of the flow:

$$\log(p(\boldsymbol{\theta}(t_1))) = \log(p(\boldsymbol{\theta}(t_0))) - \int_{t_0}^{t_1} dt \text{Tr}[J_f(\boldsymbol{\theta}(t))]. \quad (18.3)$$

Assuming a non-stiff ODE, the integration can be performed rapidly with state-of-the-art ODE-solvers, which in our case is MALI [247]. In addition, we have to solve a trace instead of a determinant, which reduces the complexity, going from $\mathcal{O}(D^3)$ to at most $\mathcal{O}(D^2)$ with D being the dimensionality of the posterior space, and speeds-up the computation. Moreover, using continuous normalizing flows removes the need to use coupling layers between transformations. Instead, all parameter dimensions can be dependent on each other throughout the flow. Combining the continuous and conditional flows leads to continuous conditional normalizing flows, where the conditional consists of the GW data d and the time t .

We also need a better data representation than the raw strain to train and analyze the data. Therefore, we follow a similar approach as in [294–298], using an SVD decomposition (see Sec. 6.4.2) as summary statistics, reducing the dimension and the noise content of the data while retaining at least 99% of the original signal. Each of the 256 generated basis vectors is used as a kernel in 1D convolutions used as an initial layer in a residual CNN (see Sec. 6.2 and in particular 6.2.2), enabling one to capture the time variance of the signal. Therefore, we do not need to use a Gibbs sampler to estimate the time of the signal as is done in [296–298], and can sample over time like any other variable which allows us to retain the likelihood of the samples. The flow itself is represented by two multi-layer perceptrons with 3 hidden layers of 512 units. Furthermore, we also use a different representation for the angles. Instead of directly using their values, we project them onto a sphere for the sky location and a circle for the other angles. This makes for a better-posed domain for these angles, plays on the interpolation capacities of the network, and makes the training step easier.

In the end, our framework combines data representation as a hybrid between classical SVD and CNN, followed by the continuous normalizing flow network. Fig. 18.1 gives a representation of our analysis framework. It is worth noting that our entire framework is relatively small compared to the ones presented in [296]. It enables the network to run on lower-end GPUs but also means it could be limited in its capacity to model the problem.

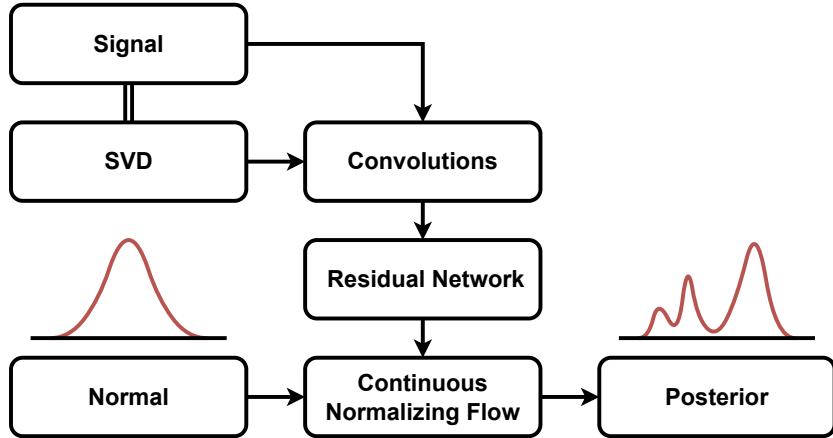


Figure 18.1: Representation of our analysis framework. It is made of a pre-processing part where we build an SVD basis to filter the data, followed by a normalizing-flow-based neural network.

18.2 Data and Setup

To test our framework’s capacity to deal with overlapping signals, we start with a simplified setup. We consider a network made of the two LIGO and the Virgo detectors, at design sensitivity [25, 26] with a lower sensitive frequency of 20 Hz. We generate stationary Gaussian noise from their PSD and inject two precessing BBH signals generated with the *IMRPhenomPv2* waveform [104]. Our data frames have an 8 s duration and are whitened before the SVD decomposition is performed. The chirp mass (Eq. (2.50)) is sampled from a uniform distribution between $10 M_{\odot}$ and $100 M_{\odot}$, and the mass ratio ($q = m_2/m_1$) from a uniform distribution between 0.125 and 1. We constrain the individual component masses to be between $5 M_{\odot}$ and $100 M_{\odot}$. Therefore, our ML approach is trained and usable on a larger range of BBH events compared to the ranges used in Chapter 17. During the data generation, the luminosity distance is kept fixed. Afterward, it is rescaled to result in a network SNR value taken randomly between 10 and 50, sampled from a beta distribution with a central value of 20. The time of coalescence for the two events is set randomly around a time of reference, with $t_c \in [t_{\text{ref}} - 0.05, t_{\text{ref}} + 0.05]$ s, ensuring that the two BBH merge in the high bias regime [360]. The other parameters are drawn from their usual domain. Table 18.1 gives an overview of the parameters and the function from which they are sampled.

During the training, we continuously generate the data by sampling the prior distributions for the events and making a new noise realization for each data frame. The training is stopped when the network has converged and before over-fitting occurs. Our model trained for about 12 days on a single *Nvidia GeForce GTX 1080* GPU.

Parameter	Function
Chirp mass (\mathcal{M}_c)	$\mathcal{U}(10, 100)$ M_\odot
Mass ratio (q)	$\mathcal{U}(0.125, 1)$
Component masses ($m_i, i = \{1, 2\}$)	Constrained in [5, 100] M_\odot
Luminosity distance (d_L)	Rescaled to follow SNR
SNR	$\mathcal{B}(10, 50)$
Coalescence time (t_c)	$\mathcal{U}(t_{\text{ref}} - 0.05, t_{\text{ref}} + 0.05)$ s
Spin amplitudes ($a_i, i = \{1, 2\}$)	$\mathcal{U}(0, 1)$
Spin tilt angles ($\theta_i, i = \{1, 2\}$)	Uniform in sine
Spin vector azimuthal angle ($\Delta\phi$)	$\mathcal{U}(0, 2\pi)$
Spin precession angle (ϕ_{JL})	$\mathcal{U}(0, 2\pi)$
Inclination angle (ι)	Uniform in sine
Wave polarization (ψ)	$\mathcal{U}(0, \pi)$
Phase of coalescence (ϕ_c)	$\mathcal{U}(0, 2\pi)$
Right ascension (α)	$\mathcal{U}(0, 2\pi)$
Declination (δ)	Uniform in cosine

Table 18.1: Summary of the parameters considered for the BBH generation and the functions used to sample the parameter values.

18.3 Results

First, we show the corner plots recovered for the masses and sky location of the two events in a pair in Fig. 18.2. These are representative of our results. One can see that the injected values are within the 90% confidence interval. This is the case for most events, regardless of the relative difference in arrival time or the SNR ratio between the two.

To demonstrate the method’s reliability, P-P plots for the two signals recovery are shown in Fig. 18.3. The P-P plot is constructed by sampling the posteriors for 1000 overlapped event pairs with parameters drawn from the distributions detailed in Table. 18.1. We then compute in which percentile of the distribution the injected value lies. If everything goes as expected, the cumulative density functions should align along the diagonal. One can observe that this is the case for our network. Comparing this to the single signal results given in Ref. [296], there is a broadening of the shell around the diagonal, showing more variability in the signal recovery. This means that our inference is less accurate than for single signals. Possible origins are the degenerate posteriors, increased complexity of the problem, and the reduced size of our network. In addition, an increased variability has been noted when going from single parameter estimation to

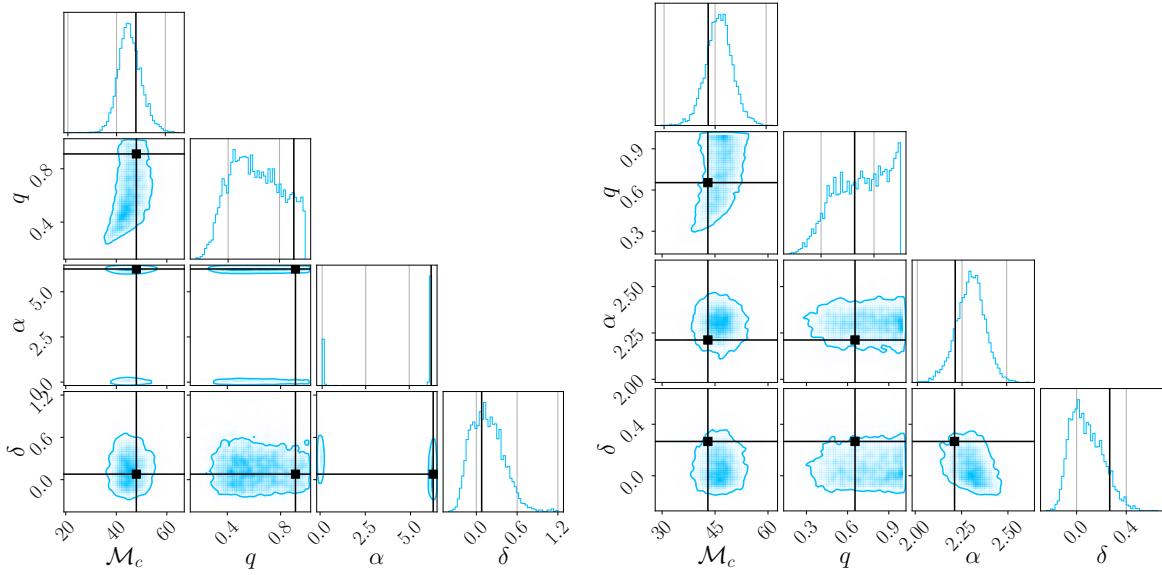


Figure 18.2: Example recovery for the masses and sky location of two overlapping BBHs with our ML approach (event A is left and event B right). The recovered values are well within the 90% confidence interval.

joint parameter estimation methods in Bayesian approaches [426].

A trick used to alleviate the problem of degeneracies is to time-order the samples. Indeed, for two BBHs, the likelihood is symmetric in the two events. Therefore, the posteriors can get bimodal [426]. While our training method formally labels one event as A and the other event as B, when the characteristics of the events are close, we may get somewhat bimodal posteriors. This probably also contributes to the higher variability of the P-P plot.

Since parameter estimation for overlapping signals is still an active field of research, it is difficult to compare with traditional methods. While methods have been developed in Ref. [426] and Chapter 17, they are not yet fully stable and take a long time to analyze a BBH system. Therefore, making a statistically significant study comparing the two approaches seems premature at this stage. However, to have some sense of the performances of our network compared to traditional methods, we consider 15 injections complying with our network’s setup and analyze them with the framework presented in [426]. Using these analyses, we can already identify some trends between the two pipelines. The first is that our ML framework typically has broader posteriors than the Bayesian approach. As mentioned in Ref. [426], the standard joint parameter estimation approach can sometimes get overconfident –see Ref. [426] or Chapter 17 for a discussion on the Bayesian algorithm–, where the recovered injected value lies outside the 90% confidence interval. Our method is not confronted with this bottleneck as the broader

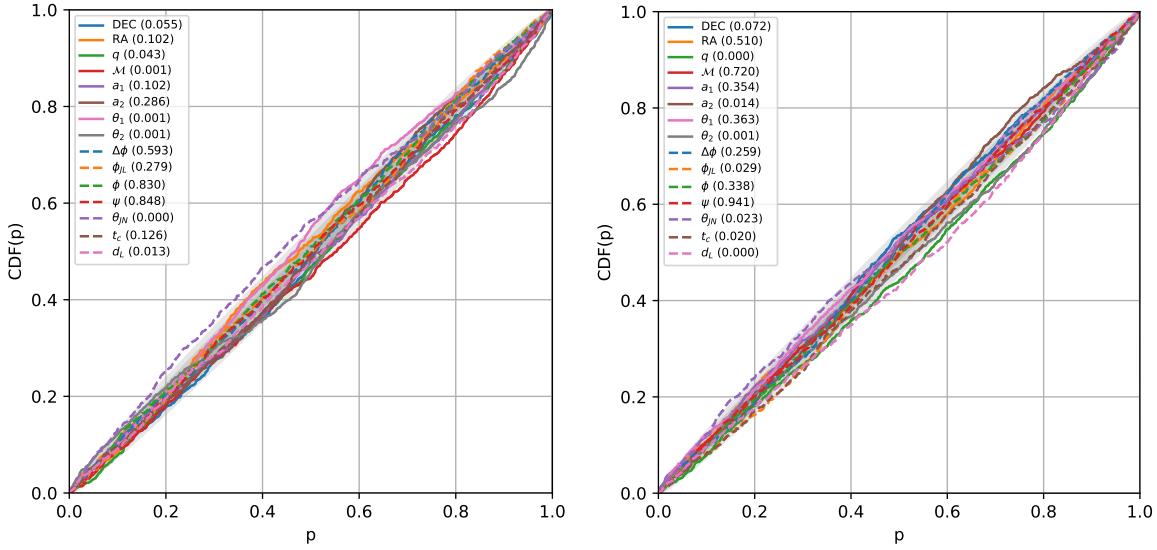


Figure 18.3: P-P plots for the recovered parameters for the two events in the data. In both cases, the lines align along the diagonal, showing that our method is trustworthy. The larger spread can be due to bimodalities in the data, the increased complexity of the problem, and the reduced size of our network. The legend indicates which line corresponds to which parameters. The values between the brackets correspond to the KS test statistic.

posterior encapsulates the injected value². Fig. 18.4 illustrates the two representative situations: one where the Bayesian approach finds the event correctly, and one where we see that our ML approach covers the injected values while it does not for the standard approach. It should be noted that an offset in the posterior, similar to the one noted in Ref. [426], can exist in our method and would not be seen because of the broad posteriors. However, because we are using the forward KL divergence, we expect the posteriors to have some support for the injected values. The origin of the larger posterior, which is not observed in the SPE ML-based methods, is probably due to the increased complexity of the problem combined with our network’s small size. One possible avenue is applying importance sampling (see Sec. 6.4.3) after the normalizing flow [298, 300]. This would increase the computational time, but the time needed to go from events to samples would still be well below the time taken by the traditional methods. However, such methods can be tricky, and additional modifications to our network could be needed.

Finally, an important advantage of our method is its speed. After training, it can analyze two overlapping BBH signals in about a second compared with $\mathcal{O}(20)$ days reported in [426]. While it is difficult to estimate the time gain for other types of signals, such as BNSs or NSBHs, we can expect the inference time after training not to be significantly larger than for BBHs for the ML approach while it significantly increases for Bayesian methods. Since computation time is a crucial aspect of studies in the 3G era,

²In our 15 injections, we find 4 for which the Bayesian approach is overconfident. This is higher than values reported in [426] and can be related to the closer merger times we are considering.

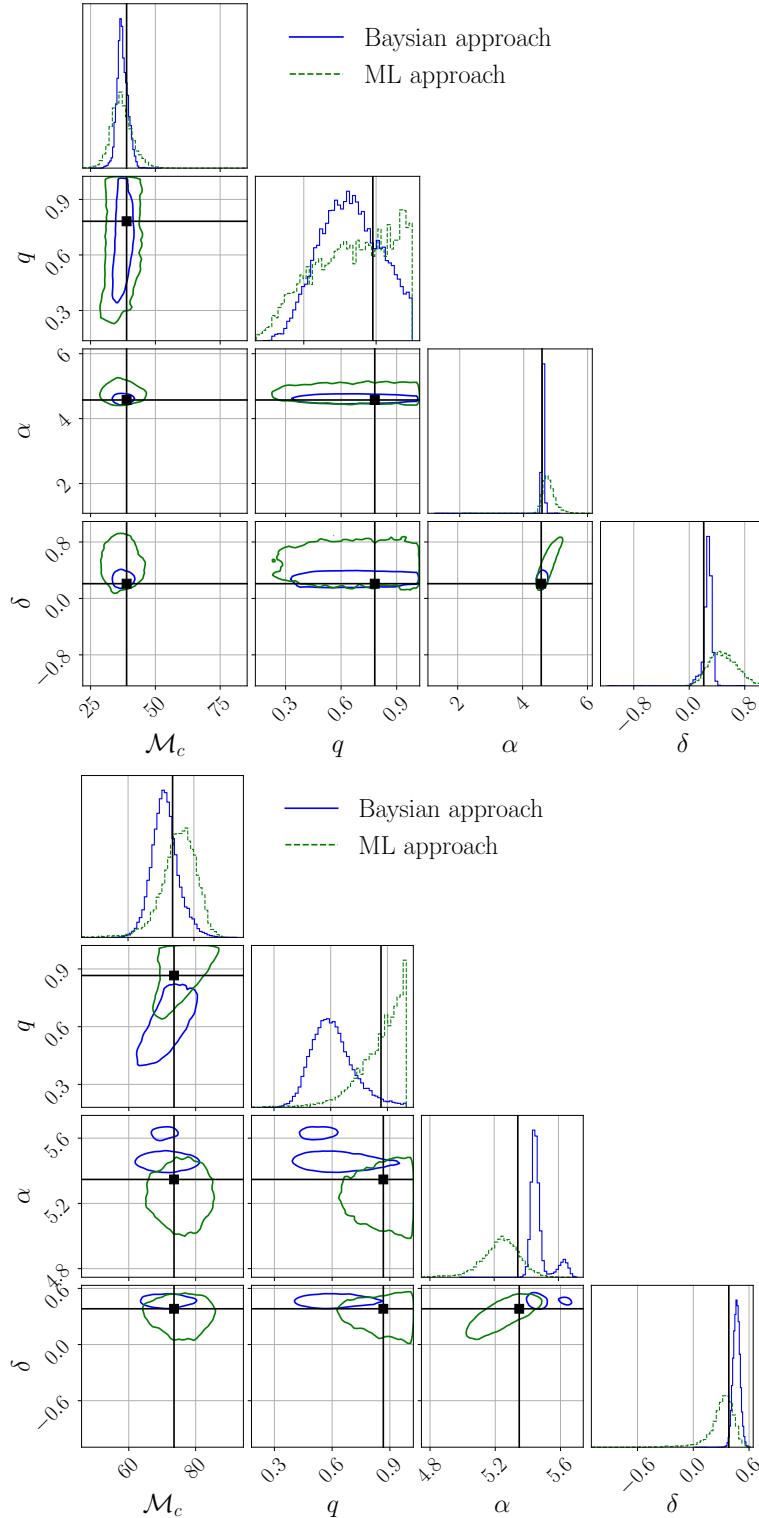


Figure 18.4: Comparison between our approach and the one from [426] for two separate events and for the chirp mass, mass ratio, right ascension, and declination. The injected values are given by the black lines. For the event in the top panel, the true value is encapsulated by the posteriors of both methods, for the bottom event this is only the case for our method. This is representative of the cases encountered in our comparisons: our posteriors are generally broader but encapsulate the injected value within the 90% confidence interval. The posteriors could be narrowed by applying importance sampling to the samples given by our network.

ML approaches seem to be more suited to study realistic scenarios for these detectors.

18.4 Conclusions and Perspectives

In this work, we have presented a proof-of-concept ML-based method to analyze overlapping BBH signals. We focused on a 2G detector scenario with the two LIGO detectors and the Virgo detector at design sensitivity, with a lower frequency cutoff of 20 Hz. Our approach is based on continuous normalizing flows.

While also using normalizing flows, as in [294–298], we bring extra modifications that seem to help in the inference task. We represent the data through a mixture of SVD and convolutions, enabling us to sample directly over the events’ arrival time, removing the need to use additional Gibbs sampling steps over that parameter and retaining the ability to access the likelihood of a sample. We also move to continuous conditional normalizing flows, reducing the computational cost of the method as we need to compute a trace instead of a determinant when going from one step to the other in the transformation. Finally, we also use a particular representation of the angles, projecting them onto circles (for the phase, the polarization, ...) and spheres (for the sky location). We believe that these modifications make our network more flexible, enabling it to deal with overlapping signals even in a reduced form.

With this simplified setup, we showed that our approach is reliable, with posteriors consistent with the injected values. Our method takes about one week to train on a single GPU. After that, it only takes about a second to analyze two overlapped BBHs. While, in reality, other types of CBC mergers can happen, their inference after training should not be significantly longer than for BBHs. We also compared our ML method with standard Bayesian methods for overlapping signals. While our scheme leads to wider posteriors, it can correctly recover the injected values, even when the Bayesian approach gets overconfident and misses the injection. In the future, we could possibly correct for the widened posteriors using importance sampling.

Our method’s combined reliability and speed show that ML is a viable approach to analyzing CBC mergers in the 3G era. More interestingly, it would even be possible without needing to account for the future development of more powerful computational means and could soon enable some more in-depth studies for ET and CE. For example, once trained for all possible BBH systems, it could help with predictions for our ability to reconstruct the BBH mass function in the 3G era.

Still, one should note that additional improvements are needed before using our method in realistic 3G scenarios. One would first need to change our setup to the 3G detectors, to account for a lower frequency cut-off and extreme SNRs. In addition, we should account for a wider range of objects. One should account for HOMs and eccentricity that could play a crucial role in the 3G era. Additionally, we need to account for the change in noise realization from one event to the other. Some of these steps, such as training

the NN for the 3G detectors instead of the 2G ones, should be relatively easy. Others are more complex, as it is hard to perform parameter inference for long-lasting mergers due to the computational burden. So, additional developments in parameter estimation using ML would be required to get to the realistic 3G scenario. For overlapping signals, one would also benefit from developments in the standard Bayesian analysis methods. For example, knowing how to deal with the noise characterization or other types of potentially detectable events.

In the end, there is still work to do before ML can be used in realistic 3G scenarios. However, we believe this work shows it is an interesting avenue and could be practical on a relatively short time scale.

Part V

CONCLUDING REMARKS AND
FUTURE PERSPECTIVES

Since its formal study by Newton in 1687, our understanding of gravity has gone a long way, undergoing some major revolutions. The biggest one was Einstein's general relativity (see Chapter 2). His geometrical description of spacetime, where massive objects deform its fabric, in turn leading to the objects' motions, explained some shortcomings of Newtonian gravity. Additionally, it made some interesting predictions. Two of these are gravitational lensing and gravitational waves. At first, the latter were believed to be unobservable because of their faintness. The presence of denser objects in the Universe, such as neutron stars and black holes, made their detection possible. After years of development to reduce the noise sources in the detectors, model as accurately as possible signals from merging compact binaries (see Chapter 3), and using adapted search methods (see Chapter 4), the first gravitational-wave detection was made in 2015. The observed signal corresponded to two merging black holes and was seen by the two LIGO detectors. In 2017, the Virgo detector joined the network of ground-based interferometers, making for the first joint binary black hole observation. It was followed rapidly by the discovery of GW170817, the first observation of a gravitational-wave signal from a binary neutron star merger. Its detection was crucial as the collision was also seen in electromagnetic bands, confirming the nature of the objects and pushing our understanding of such objects further. Since then, more gravitational-wave signals have been detected, and close to 100 compact binary mergers were reported after the last LIGO-Virgo-KAGRA observation run.

Other predictions made by general relativity have been verified by now. This is the case for gravitational lensing (see Chapter 5), where a wave gets deflected by a massive object along its travel path. For light, it was confirmed by Eddington in 1919 but has been observed many times since. Nowadays, it is routinely used to study the Universe. For example, it enables one to observe distant objects or search for exoplanets. Moreover, gravitational lensing, as described by general relativity, should also lead to lensed gravitational waves (see Chapter 5 and Part II). Depending on the regime, one would observe (i) distorted waveforms or (ii) multiple repeated events with the same frequency evolution and separated in time. This has not been observed yet, but the detection probability increases with each detector upgrade. The predicted detection rate for the coming years makes the observation of lensed gravitational-wave signals a real possibility. Many efforts to develop search techniques and science applications specific to lensed gravitational waves have been developed over the last years, including works presented in this thesis (see Part II).

Whilst successful in the detection and characterization of gravitational-wave signals in the past years, current analysis methods will face issues in the coming years. For example, with future detector upgrades and new detectors joining the ground-based interferometer network, the detection rate will increase. This requires adapted techniques able to keep up the pace. In particular, a larger number of detections mean more opportunities for multi-messenger observations. To extract the maximum information possible in all the channels, one needs to observe the event in the electromagnetic band at the time of or even before the merger (see Part III). Another problem arising from the improved

sensitivity is overlapping signals (see Part IV). For third-generation detectors, such as Einstein Telescope and Cosmic Explorer, the observed compact binary coalescence signals will be so numerous that they start overlapping, meaning that several of them are in band simultaneously. This can lead to issues and might require adapted techniques.

In parallel with technological development on the gravitational-wave side, other fields have also been improving. This is the case for machine learning, now mature enough to be used for gravitational-wave data analysis (see Chapter 6). Owing to its speed and flexibility after training, it is an interesting avenue to solve some of the issues present in gravitational-wave data analysis, such as pre-merger alert, rapid parameter estimation, and future detector studies.

Gravitational-Wave Lensing

In Chapter 7, we introduced several methods to do Bayesian analysis for strong lensing. However, because of the number of pairs one needs to analyze, one requires faster and more precise methods. Therefore, we introduced GOLUM (Gravitational-wave analysis Of Lensed and Unlensed waveform Models, see Chapter 8), a fast and precise method to search for and analyze strongly-lensed gravitational waves. It relies on recasting the joint evidence as a conditioned one, where we use the posteriors obtained when analyzing the first image as a prior for the second image. This framework is also helpful for the development of lensing studies where many injections are required. Under the lensed hypothesis, GOLUM is theoretically equivalent to joint parameter estimation.

We continued by showing how one could find smoking-gun evidence for strong lensing in Chapter 9. One of the main issues in confidently detecting strong lensing is the associated high false-alarm probability. Therefore, finding particular characteristics not easily mimicked by unlensed pairs is crucial. This is the case for the Morse phase for type II images. However, it has been shown in past studies that for single images, one requires significant higher-order mode contributions to have an observable signature. In our work, we showed that for a lensed event pair containing a type II image, the signature is visible at lower total higher-order mode content compared to the single-image case. So, we would see evidence for lensing faster. Combined with lens models, it could be used to decrease the false-alarm probability further in the future. Conversely, lensing also helps the study of the event and its higher-order mode content, enhancing various applications when such events are observed, such as tests of general relativity, for example.

Even if lensing searches were not yet successful, it remains crucial to be ready for the first detection. Therefore, in Chapter 10, we studied a way to decrease the false-alarm risk. As shown in previous work and confirmed by ours, including a lens model in the lensing searches enables one to diminish the false-alarm probability for galaxy lenses significantly. Our work also looked at the impact of errors in the lens model. We have shown that small systematics in the lens model do not hinder the benefits of adding a model. Moreover, there is no major disadvantage in using only the time delay, dropping the effect of the relative magnification, which is more uncertain. On the other hand,

if one uses a completely wrong model, for example, a galaxy-cluster lens model for a galaxy-lensed event, strong lensing can be missed. Based on this, we suggested a strategy to search for strongly-lensed events with reduced false-alarm risks: first compute the model-independent coherence ratio, then correct it for several lens models. If the event is genuinely lensed, it should keep a high coherence ratio for the lens models resembling each other. The safest way to know whether the event is lensed or not, in this case, is to use an unlensed background. One simulates a large population of unlensed events and computes the detection statistic for pairs made of these events. Then, the detection statistic for the lensed candidate is compared with the values found for the unlensed background.

Gravitational-wave lensing is a relatively new field in data analysis. Therefore, we believe there is still room for improvement in the coming years. Additionally, the perspective of a first detection is exciting. Lensed events allow probing the Universe differently. More searches are needed to reveal the potential of such events in the future. However, to exploit them, we also need to identify them confidently. Therefore, more research is required to understand the false-alarm risk and how to decrease it. Some concrete next steps are studying the impact of triple and quadruple images, looking at the impact of sub-threshold events, and checking the impact of the inclusion of the Morse factor information in lens models.

Machine Learning as Tool for Gravitational-Wave Data Analysis

The first multi-messenger observation led to a massive influx of information, motivating the community to observe more such events. However, since GW170817, no other multi-messenger event happened. Nevertheless, to make more such observations, dedicated search strategies – able to target binary neutron star events better and earlier – should be designed. Traditional methods can be used, but they are still limited as they are relatively slow and not very precise. Therefore, in Chapter 12, we introduced a machine learning-based approach to issue alerts for inspiraling binary neutron stars. In this proof of concept, we used a convolutional neural network to detect an incoming binary neutron star coalescence before its merger. We showed that our network can identify events if they have a sufficiently high signal-to-noise ratio.

Prompted by our first study, we continued the development of our early-alert system in Chapter 13. Using an improved training strategy (namely, curriculum learning on the signals' maximum frequencies), we could reduce the maximum frequency required for detection. We also tested our network on more realistic noise scenarios and showed that it could detect some events in an O4 observation scenario. Still, the network had a relatively high false-alarm rate. This can be reduced by requiring several triggers in a row, but it would increase the latency for detection. Further upgrades are also needed before using our network in low-latency online searches. Still, machine learning techniques are a promising avenue for early-alert systems.

Complementary to alerts, it is also important to develop fast and accurate sky maps

for mergers. Parameter inference for binary coalescences can be sped up with machine learning. These techniques often have reliability issues as they are limited to their training domain. Additionally, people often complain about their black-box nature. In Chapter 14, we addressed some of these concerns by setting up a hybrid sky location and mass determination system. As a first step, it used neural networks to have a base distribution for the component masses and the sky location. Then, we applied importance sampling, using the usual gravitational-wave likelihood to correct the posteriors and obtain optimal ones. This method is generally reliable but struggles somewhat with low signal-to-noise ratio signals. Further upgrades are needed to get sky maps even closer to the real ones. For example, the reflection of the sky position needs to be added to our machine learning proposal distribution. This method could be used in the future to get rapid characteristics of the events, which can then be issued to astronomers.

Combining this technique with the binary neutron star early-alert system would be valuable for future multi-messenger searches.

Third Generation Detectors: Interesting Science, Tough Analyses

Third-generation ground-based detectors will have a vastly improved detection rate for compact binary coalescences. Therefore, using a detector network made of Einstein Telescope and Cosmic Explorer, we turned to the study of overlapping signals in Chapter 16. Since third-generation detectors lead to many more detections and longer-duration signals, their overlap seems probable. First, we assessed the extent of the overlaps by simulating one year of data for a third-generation detector network. We observed that tens of thousands of mergers happen. Tens of compact binary coalescence can happen throughout a binary neutron star, and it will be common to have two signals ending within the same second. Based on this, we study the impact of overlapping signals on data analysis. Therefore, we considered the overlap of two binary black hole signals, two binary neutron star signals, or one binary black hole and one binary neutron star signal. We overlapped them with a zero or two seconds shift between their coalescence times. For the overlap of a binary neutron star and a binary black hole signal, the parameter estimation for the neutron stars was hardly affected by the overlap, regardless of the difference in merger times. Conversely, for binary black holes, the recovery was affected. For a high-mass binary black hole, the sampler was unable to find the signal in the data. For a lower-mass signal, we retained some measurability but observed some deviations compared to the case without an overlapping binary neutron star. When the time shift is bigger, the biases mostly disappeared. For two binary black holes with sufficiently dissimilar masses, the parameter recovery was unbiased. For two binary neutron star signals overlapping, when they merge at the same time, the faintest signal was not recovered. The problem was solved when the merger times were different enough. In the end, this showed that, when the events' merger times are close, biases can occur, and one may need to employ adapted inference techniques.

Independently of our studies, other authors also showed that biases occur when

overlapping signals are analyzed without accounting for the other signals' presence. Therefore, it is important to introduce adapted inference techniques. Such methods were presented for the first time in Chapter 17, where we compared two approaches: hierarchical subtraction and joint parameter estimation. For the first, one analyzes the dominant signal, subtracts the result, and then infers the second one. This can lead to bias when the first signal is characterized badly. Joint parameter estimation analyzes the two signals together. It is less keen to bias but is computationally heavier due to the increased dimensionality of the parameter space. Focusing on high-mass systems due to limited computational resources, our work shows that these adapted methods can reduce the biases present in the recovered posteriors. This is crucial for realistic applications when considering third-generation detectors. However, these methods are still subject to some instabilities, and further upgrades are needed to have optimal results. Moreover, the techniques are computationally heavy, making them slow to use. With their current computation time, it would be impossible to follow the detection rate expected during the third-generation detector. Therefore, they need to be accelerated, for example, by using relative binning or adaptive frequency resolution.

Recently, parameter estimation for binary black holes using machine learning techniques has emerged. Even if it is not always entirely reliable yet, these techniques offer major speed-ups once trained. So, it is interesting to see if such methods are usable for overlapping signals. In Chapter 18, we presented a normalizing flow-based approach designed to analyze overlapped binary black hole signals. In this proof-of-concept work, we introduced several upgrades to the existing machine learning-based parameter estimation frameworks and analyzed overlapping binary black hole signals in a second-generation detector network. We showed that our network can recover the injected values reliably. Unfortunately, our posteriors were often broader than those obtained through joint Bayesian inference. Still, in the future, this could be corrected by using importance sampling. Moreover, our network is about half the size of those used for single event parameter estimation, while the parameter space is twice as big. Therefore, increasing the network's complexity should improve the final precision of the recovered parameters while retaining a low computational cost compared to Bayesian approaches. However, it could mean our network cannot run on low-end GPUs anymore. Our study shows that machine learning is a viable alternative for studying signals in the third-generation era, but some progress still needs to be made to deal with the duration and variety of the signals.

Whether with traditional Bayesian methods or machine learning, developing methods to analyze signals in the third-generation era is crucial. In addition to close-by mergers, other problems remain to be solved. This is the case, for example, for noise characterization or analyzing an event plagued with tens of mergers occurring over its duration. These issues should be addressed if we want to extract the full potential from these detectors. To be ready when the third-generation detectors come online, mock data challenges and science cases should be pursued. They need to be as realistic as possible, meaning the required analysis tool should be developed as early as possible. Once these

bottlenecks are overcome, third-generation detectors should lead to a major increase in our understanding of the Universe, thanks to their remarkable observational capacities.

Final Words

We are living in exciting times for gravitational-wave astronomy as current detectors are getting closer and closer to their design potential. This should unravel new phenomena, such as lensing. In this thesis, we presented several studies developing frameworks necessary to search for and characterize strongly lensed gravitational-wave events. We also showed some interesting applications proper to gravitational-wave lensing. Additionally, new interferometers are being designed and commissioned. Their construction requires major technological development. Their exploitation requires the introduction of novel and fast data-analysis techniques as the observed merger rate increases. In this work, we have shown that overlapping signals will happen in third-generation detector networks, and can lead to biased posteriors for the compact binary coalescence parameters if not accounted for. We also performed the first joint parameter analysis for overlapping signals, opening the way to the development of more realistic analysis frameworks for the third-generation detector era. Once mastered, these detectors should push human knowledge further, beyond previously-explored horizons. In parallel, new research fields are developing, such as machine learning and quantum computing. To extract the maximum potential of our detectors, looking at and using the developments done in other research fields is crucial. Several works presented in this thesis are first steps toward linking gravitational-wave data analysis and machine learning to prepare for the future. For example, we showed how valuable machine learning could be for multi-messenger astronomy by developing an early-alert system for inspiraling binary neutron stars. Besides, we also illustrated its application for the analysis of overlapping binary black holes, drastically reducing the analysis time compared to Bayesian methods, and opening the door to more realistic studies for third-generation detectors.

PUBLIC SUMMARY

From Apple to Modern Description of Gravity

Newton's apple is a story often remembered by people as the beginning of the study of stellar and planetary motions. However, this is not true. Already in its early stages, humankind was curious about the shiny little dots in the sky at night. For some, they were manifestations of the gods. For others, they were revealing the beauty of the Universe. Already during Antiquity – from the Greeks to the Babylonians, to the Egyptians, to the Maya, to the Chinese –, one can find traces of people studying gravity and stellar objects. Their basic understanding came from observation and interpretation, often cast as philosophical arguments. Nevertheless, this is where things started.

After the Middle Age, when people re-discovered antique works, a new interest in studying the Universe sparked among scholars, making for several new propositions, even leading to conflict with the Church as people moved away from purely biblical interpretations to observation-based ones. In particular, we went from a geocentric to a heliocentric description of the Solar System, where the Earth is only one of the Sun's satellites. Several now-famous figures have helped in this development. Post-antiquity, Copernicus was the first to suggest the idea. It was supported further by Galileo, who defended it based on the observations he made – using one of the first telescopes – of Jupiter's moons. So, the Earth would be a simple satellite from the Sun. Based on this, Kepler issues his laws of planetary motions. These empirical laws are the first laws describing the planets' movements in the Solar System. One shortcoming in Kepler's theory is that the cause of this motion is not explained. Newton found the explanation by introducing his laws of motion, where he introduces the notion of forces and describes the gravitational force: every massive object attracts other massive ones. This force and the laws of motion enable one to explain the planets' orbits.

Years later, Newton's depiction of gravity was challenged by more precise observations, as it could not explain the advance of the perihelion – the point of closest approach to the Sun – in Mercury's orbit. This was found by Le Verrier, who suggested the existence of an extra planet between the Sun and Mercury explaining the observed discrepancy. This planet has never been observed, but another explanation was found, thanks to a new theory: general relativity, derived by Einstein at the beginning of the 20th century. At first, it was received with some skepticism as it has a strange entanglement between a geometric description of spacetime and forces. However, the theory rapidly gained popularity by explaining observations not justified by Newtonian gravity (such as Mercury's orbit) and recovering the same result in cases where the latter was working. Moreover, the theory led to new predictions. Observing those would validate it even further. Some of these predictions, considered in this thesis, are gravitational lensing – when a massive object bends the path of a light ray from source

to observe –, and gravitational waves – where a cataclysmic phenomenon leads to a perturbation in the spacetime fabric propagating as a wave through the Universe. Such predictions have been verified, making general relativity a successful theory.

Ripples in Spacetime

Gravitational waves are often called ripples in spacetime by analogy with waves propagating on a pond when one throws a stone in it. However, the mechanism is a bit more involved. Because spacetime is rigid, one needs massive amounts of energy and momentum to be converted into gravitational waves for them to be detectable. Even so, their direct detection requires measuring relative length changes of about one thousand of an atomic nucleus.

Gravitational-wave emission was predicted by Einstein, but he thought it would never be observed because of its faintness. Actually, he was right in the scenario he considered: the collision of stars. Years later, the discovery of neutron stars and black holes changed the game. These extremely massive objects can lead to the emission of gravitational waves with a potentially detectable effect. Nevertheless, to achieve their direct observation, we needed to overcome many technological challenges. The first hint of their existence was obtained through the observation of the Hulse-Taylor binary pulsar, made of a neutron star and a pulsar. Such a system can be seen with a telescope, and its monitoring showed a decaying orbit consistent with an energy loss through gravitational waves, compatible with the prediction made by general relativity.

The development of gravitational-wave detectors – aiming for direct detection – started in the 1960s. Years of theoretical and technological developments lead to the construction of the LIGO (Laser Interferometer Gravitational-wave Observatory) detector. More improvements were required to bring it to a stage where it could detect gravitational waves emitted by mergers made of the densest objects in the Universe. On the 14th of September 2015, the first-ever direct detection was achieved when the gravitational-wave event GW150914 was recorded. It corresponded to the encounter of two black holes with masses of about 36 and 29 solar masses. This observation is a crucial step in astronomy as it confirms the existence of black holes, but also that such objects can form, combine, and merge in a time shorter than the age of the Universe.

More binary black holes were observed, and in August 2017, a new observatory, Virgo, joined the network. The presence of a third detector helped in determining the events' sky location. Soon after Virgo joined, another breakthrough happened: the observation of GW170817, corresponding to the encounter of two neutron stars. Since those objects are made of matter and do not have a horizon, they also lead to electromagnetic emissions. Those were observed independently by several facilities, linking gravitational-wave and electromagnetic data. Observing such a phenomenon led to new ways to understand how dense elements (such as gold) form in the Universe, test general relativity, and understand the neutron star equation of state.

Since then, a third observation run took place. During it, we observed more binary black holes. We also saw some neutron star binaries but did not identify an electromagnetic counterpart. Still, some new phenomena were seen. For example, we observed the first two mergers of a neutron star and a black hole, found a binary black hole with theoretically unexpected component masses, and detected higher-order modes³ for the first time. With about one hundred binary mergers detected, we can reconstruct the mass distributions of black holes and neutron stars, study their merger rates, and perform tests of general relativity in a regime that is not obtainable on Earth, among other things.

Gravitational-Wave Lensing

The deflection of a wave by a massive object is also a direct consequence of general relativity. It was confirmed by Eddington for light when he observed the deflection of starlight by the Sun during an eclipse. Nowadays, the lensing of light is a tool commonly used in astronomy. Interestingly, massive objects can also deflect gravitational waves. If the lens is dense enough, we can distinguish several “images” seen as (pairs, triples, or quadruplets, … of) repeated gravitational-wave signals with the same frequency evolution. This phenomenon – called strong lensing – has not been observed yet, but it is predicted at a non-negligible rate in the coming years. Therefore, we are already searching for it.

Typically, when looking for strongly-lensed gravitational waves, one looks for event pairs with the same intrinsic parameters and sky locations, linked by an overall phase shift, a magnification, and a time delay. Therefore, one needs to analyze all the pairs of events in the data, which is a rapidly growing number as the detection rate increases. Moreover, if the search technique is not precise enough, we can have false lensing claims as parameters seem to match even if they do not. Therefore, in this thesis, we introduced a new search method combining speed and precision to search for strong lensing, effectively filling the gap between pre-existing pipelines, where some were precise but slow and others fast but not precise. We also showed how this framework, called **GOLUM** (for *Gravitational-wave analysis Of Lensed and Unlensed waveform Models*) reduces the computational cost for joint parameter estimation to that of a usual binary black hole analysis.

Strong lensing is also associated with a high false-alarm risk. Even with precise methods, we only have a finite precision and some error bounds on the inferred parameters. Therefore, we may see matching events by chance when many merger signals are present in the data. To make a confident lensing detection, one needs to alleviate this. In this

³A bit like an instrument, gravitational waves can be described with different modes (or tones). There is the dominant one, which we measure the best, but other tones are also present to make the complete symphony of the black hole signal. These additional modes are generally fainter and harder to detect.

thesis, we explored two different avenues to identify strongly-lensed images. The first consists of using higher-order modes to detect typical lensing signatures not present for unlensed events. Past studies have shown that, in some cases, the overall phase shift due to strong lensing can be unequivocally detected, provided the higher-order mode content is high. However, its detection requires particular events, unlikely for current detectors. In our work, we showed that if one is confronted with a pair of events and analyzes them jointly (as done in any strong-lensing search), this particular feature is detectable at much lower higher-order mode content. Therefore, this smoking-gun evidence for lensing could be observed in more common scenarios.

A second avenue to reduce the false-alarm probability explored in this thesis is the inclusion of lens models in the search pipelines. As said earlier, binary parameters can match by chance. However, one also has so-called lensing parameters linking the two events. For a given lens model, these parameters have expected distributions. For example, for a galaxy lens, the time delay between images should not be more than a few months and is most likely between one and a few days. The chance of unlensed event pairs having matching parameters and appropriate apparent lensing parameters is much smaller than the chance of simply having similar binary parameters. Therefore, including this in the search information is important when facing many events. Generally, search pipelines are independent of the lens model to avoid missing a lensed event not complying with the model. In this thesis, we showed how one can include such models after the search is done, avoiding the computationally-expensive procedure where the analyses have to be repeated for every model. In addition, we show a consequent reduction of the false-alarm probability when including the correct lens profile and demonstrate that small systematics in the lens model would not hinder its decrease. Conversely, using the wrong density profile could make lensing detection impossible. Based on this, we suggest a search strategy for strong lensing: start with model-independent searches to identify the most promising candidates. Then, we include an array of models in the detection statistic with redundancy between lens profiles to account for systematics. A genuinely lensed pair would keep a high detection statistic for all the lens models of the same family. While not guaranteeing detection, this is the safest way to search for strongly-lensed event pairs.

Early-Alert Systems for Binary Neutron Stars Using Machine Learning

Over the last century, gravitational-wave physics was not the only field having breakthroughs. It is also the case for machine learning, which went from being a concept to a powerful tool that is commonly used. Nowadays, it is mature enough to be used in other research fields. For gravitational-wave data analysis, one of its most desirable features is the speed after training. Indeed, one often has to use computationally expensive methods requiring a long time to converge or needing to loop over many parameter

combinations to find a trigger. This can be harmful in some cases. The computational cost is moved upstream for machine learning, during the training. Once trained, the algorithm can generally produce results within a few seconds. This is desirable, in particular in low-latency searches. These are critical for binary neutron stars, where we want to observe the electromagnetic emission associated with the merger and its early stages. Studying the event at the merger time or even before would be crucial to better understand the matter's behavior under these extreme conditions.

In this thesis, we built the first early-alert system for inspiraling binary neutron stars using machine learning. We made a convolutional neural network to trigger when it observes an inspiraling binary neutron star. In the first instance, we took a reduced time window where the signal starts at 20 Hz and split the binary neutron star into three mass categories. We trained and tested the network on design sensitivity data for a single LIGO detector and found that we can detect inspiraling binary neutron stars provided they are loud enough. In this proof-of-concept study, the setup was not entirely realistic as we did not yet change the sky location or account for the detector network. Therefore, in a second step, we improved our framework. We decreased the minimum frequency seen by our network to something around 12 Hz. In addition, we increased the window size, removing the need for mass categories. We also used a new training technique: curriculum learning, where we iteratively increased the difficulty in the data seen by the network when training. Applying all these improvements, we trained and tested our network on different noises: simulated noise corresponding to the third and fourth observation runs, and real noise recorded during the third observation run. We see that our performances slightly degraded for real, less well-behaved, data. However, we retained some detectability. Moreover, the improved training strategy makes for increased sensitivity. For an O4-like observation scenario, about 20 binary neutron star mergers could be detected over a year. Our network would issue an alert for about 3 of them, with a 90 s delay before the merger in some cases. Unfortunately, our algorithm also has a high false-alarm rate, leading to numerous spurious triggers. It can be decreased by asking for multiple triggers in a row when analyzing data. However, this solution cuts back the time between the trigger and the merger. In the end, this work shows that machine learning is a viable alternative for gravitational-wave early alert, even if further developments are required before it is practically usable.

In addition to a trigger, one needs to tell astronomers where to point their telescopes when searching for a counterpart. Therefore, it is important to have fast ways to produce a sky map for an event. This is also a place where machine learning can be helpful. However, one complaint people often have is the reliability and the dependence on the noise realization used for training. Therefore, in this thesis, we pursued a hybrid approach, combining machine learning and Bayesian statistics, to rapidly obtain the sky map and the mass distribution for a given binary black hole event. We used a neural network to make a well-posed proposal distribution for the masses and sky location. Then, we used importance sampling to transform the distributions into the final ones. In the end, our framework can produce sky maps in a few minutes, compared to days for traditional

Bayesian methods. Moreover, the number of effective importance samples gauges the quality of the proposal distribution and the neural network's behavior. However, this framework still needs some improvements. For example, our proposal distribution for the sky location does not account for a possible sky reflection. At any rate, this shows machine learning is a promising avenue for prompt localization and rapid parameter inference.

Adapting the sky location inference framework and the binary neutron star early alert system would make for the first machine learning-based early-alert system.

Third-Generation Detectors

Even before the first gravitational wave was detected, planning for the next generation of ground-based detectors had started. In essence, these detectors are bigger and more sensitive than current ones, leading to more observations of compact binary merger signals and the detection of new phenomena. Next-generation detectors are the European Einstein Telescope, generally seen as a triangular network with 10 km long arms, and the American Cosmic Explorer, retaining the L-shaped design but drastically increasing the arm lengths to 20 or 40 km. In addition to their increased size, their technologies should be improved, extending the reach for binary black holes up to the cosmic dawn, where only primordial black holes –originating from inhomogeneities in the big bang– would exist. Binary neutron stars should be observed up to a redshift of a few, making for more multi-messenger detections and a better grasp of their spatial distribution.

Because of their increased sensitivity and reach, third-generation detectors will see more signals, and signals will have a longer duration. Therefore, it is likely they start to overlap in the detection band, with several of them visible simultaneously. In this thesis, we first study how often such overlaps would happen. We simulate one year of data in a third-generation detector network. Considering only binary neutron stars and black holes, overlaps seem relatively common, with tens of merger signals happening throughout a single binary neutron star signal. Moreover, tens to thousands of merger signals end within the same second, depending on the local merger rate considered. So, overlapping signals will occur in the third-generation detectors era. Since we know they will happen, we look at their impact on the data analysis. To that end, we looked at the overlap of two binary black hole signals, two binary neutron star signals, and one binary black hole and one binary neutron star signal merging concurrently or two seconds apart. When neglecting one of the two signals, we find that bias can occur in different situations. Biases were observed in the parameters recovered for the binary black hole signal when overlapped with a neutron star signal. In particular, we were not able to find the signal at all for the heaviest binary black hole considered. The binary neutron star parameter recovery was not affected by the overlap. When two binary neutron star signals were overlapped, we were unable to recover the characteristics of the faintest one. The other signal was not affected by the overlap. In the other cases, the parameter recovery was

not affected by the overlap. However, current parameter estimation methods may work in some situations but not others. Therefore, adapted tools need to be developed.

Soon after our study on biases, other studies confirmed the possibility of observing distortions. In particular, for binary black holes with close characteristics, errors in the recovery can also occur when they have close merger times. In addition, the bias depends on the exact characteristics of the signal. Therefore, it is of interest to develop adapted analysis tools. In this thesis, we presented two possible Bayesian analysis methods: (i) hierarchical subtraction, where we analyze the dominant signal, subtract it and then analyze the second, and (ii) joint parameter estimation, where we fit the two signals at the same time. The second approach is more precise as it accounts correctly for the noise characteristics for the two signals, but it is computationally heavier due to a larger parameter space. We compared the results of these two methods with those obtained when analyzing the signals without overlap. Here, due to computational restrictions, we focus on higher-mass binary black holes merging 0.1 s apart. We found that, as expected, we recurrently have biases when analyzing the dominant signal for hierarchical subtraction. However, this bias is not always echoed in the second signal. Therefore, we tried applying a third step where we subtract the second signal and analyze the first one. We did not observe significant improvement in the recovered posteriors. For joint parameter estimation, we found more accurate results than for hierarchical subtraction. However, they are not equivalent to single parameter estimation due to the increased complexity of the problem and degeneracies. Still, with some improvements, one could get good results with joint parameter estimation and eliminate the remaining instabilities. This study performed the first-ever joint analysis of the two gravitational wave signals and paves the way for future developments.

One of the drawbacks of the Bayesian methods to analyze overlapping signals is their speed. Currently, they would be unable to follow the detection rate of third-generation detectors. Therefore, in this thesis, we also explored an alternative analysis method, using normalizing flows, to jointly analyze overlapping binary black hole signals. With this machine learning-based method, once the network is trained, we would obtain posteriors for two overlapping signals in about a second, compared to 20 to 30 days for joint parameter estimation. We found that our machine-learning framework can recover the parameters of the two signals. It can do so consistently, even if more variation is observed compared to single parameter estimation using normalizing flows. When compared to traditional Bayesian methods, our network typically has broader posteriors. The probable principle cause is the smallness of the network, which is twice smaller than those used for single signal parameter estimation for a two-times larger parameter space. Rather than simply increasing the network size – meaning it cannot run on low-end GPUs anymore –, we could get more accurate posteriors by reweighting the network’s output samples to match those given by the Bayesian methods. Still, this is not easy due to the difficulty to obtain properly correlated samples. Nevertheless, this study shows machine learning is a viable avenue to study overlapping signals, even when accounting for high detection rates. Some improvements are still needed, but this technique would

also enable one to perform more realistic studies for third-generation detectors without needing to wait for future developments of more powerful tools.

In the end, third-generation detectors would be facilities helping to push our knowledge significantly forward. However, some bottlenecks on the data analysis side still exist and need to be dealt with to extract the full potential of these interferometers. In this thesis, we have studied some of these bottlenecks and suggested possible solutions.

OPENBARE SAMENVATTING

Van de appel tot de moderne beschrijving van de zwaartekracht

De appel van Newton is een bekend verhaal dat mensen vaak aanzien als het begin van de studie van stellaire en planetaire bewegingen. Dat is echter niet waar. Al in vroegere tijden was de mensheid nieuwsgierig naar de glimmende stipjes die 's nachts aan de hemel te zien waren. Voor sommigen waren het manifestaties van de goden. Voor anderen onthulden zij de schoonheid van het heelal. Vanaf de Oudheid – van de Grieken tot de Babyloniers, tot de Egyptenaren, tot de Maya's, tot de Chinezen – zijn er al sporen te vinden van mensen die zwaartekracht en stellaire objecten waarnamen. Hun begrip kwam van observaties en interpretaties, vaak in de vorm van filosofische argumenten. Dit is waar het verhaal echt begon.

Na de Middeleeuwen, toen mensen antieke werken herontdekten, ontstond er onder geleerden een nieuwe belangstelling voor het bestuderen van het heelal. Dit leidde tot verschillende nieuwe stellingen, die zelfs conflicten met de Kerk opbrachten omdat men van Bijbelse naar op waarnemingen gebaseerde interpretaties overging. In het bijzonder ging men over van een geocentrische naar een heliocentrische beschrijving van het zonnestelsel, waarbij de Aarde slechts één van de satellieten van de Zon is. Verschillende bekende figuren hebben aan deze ontwikkeling bijgedragen. Na de oudheid was Copernicus de eerste die op het idee was gekomen. Het werd verder ondersteund door Galileo, die het verdedigde op basis van de waarneming die hij – met één van de eerste telescopen – van de manen van Jupiter deed. De Aarde zou dus een eenvoudige satelliet van de Zon zijn. Op basis hiervan werkte Kepler zijn wetten van de planeetbewegingen uit. Deze empirische regels zijn de eerste wetten die de beweging van de planeten in het zonnestelsel beschrijven. Een tekortkoming in zijn theorie is dat de oorzaak van de beweging niet uitgelegd wordt. Newton vond de oplossing met zijn bewegingswetten, waarin hij de notie van kracht introduceert en de zwaartekracht beschrijft: elk massief object trekt andere massieve objecten aan. Samen maken deze kracht en de bewegingswetten het mogelijk het traject van de planeten te beschrijven.

Jaren later werd Newtons voorstelling van de zwaartekracht in twijfel getrokken door nauwkeurigere observaties. De theorie kon de vervroeging van het perihelium – het punt waar het de Zon het dichtste nadert – van Mercurius niet verklaren. Dit werd door Le Verrier ontdekt. Om dit op te lossen stelde hij het bestaan van een planeet tussen de Zon en Mercurius voor. Niemand heeft deze planeet ontdekt. Er is wel een andere verklaring gevonden dankzij een nieuwe theorie: algemene relativiteit, opgesteld door Einstein in het begin van de 20^{ste} eeuw. Eerst werd deze met scepsis ontvangen omdat er een vreemde verstrengeling is tussen de geometrische beschrijving van ruimtetijd en krachten. Nadien wint de theorie redelijk snel in populariteit omdat het observaties verklaart

die niet door Newtoniaanse zwaartekracht kunnen worden uitgelegd (zoals het traject van Mercurius) en hetzelfde oplevert in gevallen waarin de Newtoniaanse beschrijving werkt. Bovendien leidt de theorie tot nieuwe voorspellingen. Verificatie daarvan zouden algemene relativiteit nog verder valideren. Enkele van deze voorspellingen, die in deze proefschrift bestudeerd zijn, zijn lensing – wanneer een massief voorwerp het pad van een lichtstraal afbuigt – en zwaartekrachtgolven – waarbij een cataclysmisch fenomeen tot een verstoring in de ruimtetijd leidt die zich als een golf door het heelal voortplant. Deze voorspellingen werden waargenomen, wat algemene relativiteit tot een succesvolle theorie heeft gebracht.

Rimpelingen in de ruimtetijd

Zwaartekrachtgolven worden vaak rimpelingen in de ruimtetijd genoemd naar analogie met golven die zich op een vijver voortplanten wanneer men er een steen in gooit. Eigenlijk is het mechanisme iets ingewikkelder. Omdat ruimtetijd star is moeten enorme hoeveelheden energie en momentum in zwaartekrachtgolven omgezet worden voordat ze gedetecteerd kunnen worden. Zelf dan vereist hun directe observatie het vermogen om lengteveranderingen van ongeveer een duizendste van een atoom te meten.

Zwaartekrachtgolven werden al door Einstein voorspeld, maar hij dacht dat ze nooit gezien zouden kunnen worden omdat ze te zwak waren. In feite had hij gelijk in het scenario dat hij beschouwde: de botsing van sterren. Jaren later veranderde het spel met de ontdekking van neutronensterren en zwarte gaten. Deze extreem massieve objecten kunnen tot de potentieel detecteerbare emissie van zwaartekrachtgolven leiden. Om ze rechtstreeks te kunnen waarnemen moesten we echter veel technologische uitdagingen aangaan. Daardoor was de eerste aanwijzing van hun bestaan verkregen door de observatie van de Hulse-Taylor binaire pulsar, bestaande uit een neutronenster en een pulsar. Een dergelijk systeem kan met een telescoop opgevolgd worden, en de monitoring ervan toonde een afnemende baan aan die overeenkomt met een energieverlies door zwaartekrachtgolven zoals voorspeld door algemene relativiteit.

De ontwikkeling van zwaartekrachtgolfdetectoren – gericht op rechtstreekse detectie – begon in de jaren 60. Jaren van theoretische en technologische ontwikkelingen leidden tot het ontstaan van de LIGO (Laser Interferometer Gravitational-wave Observatory) detector. Meer verbeteringen waren nodig om deze detector in staat te stellen om gravitatiegolven afkomstig van de dichtstbijzijnde objecten in het heelal te detecteren. Op 14 september 2015 werd de eerste rechtstreekse observatie gedaan, toen het signaal GW150914 werd geregistreerd. Het kwam overeen met de ontmoeting van twee zwarte gaten met massa's van ongeveer 36 en 29 zonsmassa's. Deze observatie was een cruciale stap in de astronomie. Het bevestigde het bestaan van zwarte gaten, maar ook dat deze objecten zich kunnen vormen, combineren, en samensmelten in een tijd die korter is dan de leeftijd van het heelal.

Meer dubbele zwarte gaten werden waargenomen. In augustus 2017 voegde een nieuw

observatorium, Virgo, zich bij het netwerk. De aanwezigheid van een derde detector helpt in het bepalen van de location aan de hemel van de detecties. Kort na de toevoeging van Virgo vond een nieuwe doorbraak plaats: de observatie van GW170817, die overeenkomt met de zwaartekrachtgolfsignaal van twee neutronensterren. Aangezien deze objecten uit materie bestaan en geen horizon hebben, leiden ze ook tot elektromagnetische emissie. Die werd door verschillende faciliteiten onafhankelijk gevonden, waarbij zwaartekrachtgolven en elektromagnetische gegevens samengekoppeld konden worden. Dit soort gebeurtenis observeren leidt tot nieuwe manieren om te begrijpen hoe zware elementen (zoals goud) zich in het heelal vormen, algemene relativiteit te testen en de toestandsvergelijking van neutronensterren te begrijpen.

Sindsdien heeft een derde observatieronde plaatsgevonden. Daarbij hebben we nog meer binaire zwarte gaten gezien. We zagen ook enkele binaire neutronensterren, maar identificeerden geen elektromagnetische tegenhanger. Toch werden er enkele interessante gebeurtenissen geobserveerd. We hebben bijvoorbeeld de twee eerste fusies van een neutronenster met een zwart gat gezien, een paar zwarte gaten gevonden met theoretisch onverwachte hoge massa's, en hoger-orde modes⁴ voor de eerste keer waargenomen. Met ongeveer honderd gedetecteerde fusies kunnen we, onder andere, de massaverdeling van zwarte gaten en neutronensterren reconstrueren, hun fusiesnelheid bestuderen en algemene relativiteit testen in een regime dat op Aarde niet haalbaar is.

Lensing van zwaartekrachtgolven

De afbuiging van een golf door een massief voorwerp is ook een rechtstreeks gevolg van de algemene relativiteit. Het werd door Eddington voor licht bevestigd toen hij de afbuiging van sterlicht door de Zon tijdens een eclips observeerde. Tegenwoordig is lensing van licht een veelgebruikt instrument in de astronomie. Interessant is dat massieve objecten ook zwaartekrachtgolven kunnen afbuigen. Als de lens zwaar genoeg is kunnen we verschillende “beelden” als meerdere herhaalde zwaartekrachtgolfsignalen met dezelfde frequentie-evolutie zien. Dit fenomeen – die sterke lensing genoemd wordt – is nog nooit geobserveerd, maar heeft een interessante verwachte waarnemingswaarschijnlijkheid voor de komende jaren. Daarom wordt er nu al naar gezocht.

Bij het zoeken naar sterk gelensde zwaartekrachtgolven zoekt men gewoonlijk naar paren van detecties met dezelfde intrinsieke parameters en hemelpositie, verbonden door een algemene faseverschuiving, een versterking en een tijdsvertraging. Daarom moet men alle detectieparen in de gegevens analyseren, wat een snelgroeidend aantal is naarmate het aantal detecties toeneemt. Bovendien, als de zoektechniek niet precies genoeg is, kunnen we valse beweringen van lensing krijgen omdat parameters die niet kloppen samen lijken

⁴Een beetje zoals muziekinstrumenten kunnen zwaartekrachtgolven met verschillende modi (of tonen) beschreven worden. Er is een dominante toon die we het beste meten, maar er zijn ook andere tonen aanwezig om het complete symfonie van het zwaartekrachtgolfsignaal te maken. Deze bijkomende tonen zijn meestal zwakker en moeilijker te detecteren.

te passen. Daarom introduceren wij een nieuwe zoekmethode voor sterke lensing die snelheid en precisie combineert. Het vult het gat op tussen reeds bestaande algoritmes, waarvan sommige accuraat maar traag zijn en andere snel maar niet nauwkeurig zijn. We hebben ook getoond hoe dit kader, dat GOLUM (voor *Gravitational-wave analysis Of Lensed and Unlensed waveform Models*) heet, de computationale kost voor gezamenlijke parameterschatting terugbrengt tot die van een gebruikelijk zwarte gat analyse, wat de onderzoekstijd aanzienlijk verlaagt. De tocht naar sterk gelensde zwaartekrachtgolven is al een paar jaar bezig. Helaas heeft het tot nu toe nog niet tot een detectie geleid.

Sterke lensing gaat met een hoog risico op vals alarm gepaard. Zelf met accurate methoden hebben we alleen maar foutgrenzen voor de afgeleide parameters. Als de gegevens veel fusies bevatten, kunnen we daarom bij toeval overeenkomstige signalen zien. Om een betrouwbare lensdetectie te maken moet men de kans op een vals alarm verlagen. In dit proefschrift hebben we hiervoor twee mogelijkheden onderzocht. De eerste bestaat uit het gebruik van hoger-orde modes om een typische lensingsignatuur die niet aanwezig is in niet-gelensde signalen te detecteren. Voorgaande studies hebben getoond dat in sommige gevallen de algemene faseverschuiving die door sterke lensing veroorzaakt wordt gedetecteerd kan worden voor een enkel gelensd beeld, als de hoger-orde modes sterk genoeg zijn. De detectie ervan vereist echter bijzondere gebeurtenissen, die voor de huidige detectoren onwaarschijnlijk zijn. In ons werk hebben we laten zien dat als men een paar signalen gezamenlijk analyseert (zoals voor elke zoektocht naar sterke lensing gedaan is), dit specifieke kenmerk voor een lagere hoger-orde mode-inhoud geïdentificeert kan worden. Daarom zouden we deze herkenbare lensing signatuur in meer scenario's zien.

Een tweede manier om de kans op vals alarm in verband met sterke lensing te verminderen, die in dit proefschrift onderzocht wordt, is het gebruiken van lensmodellen in de zoekalgoritmes. Zoals gezegd kunnen binaire parameters toevallig overeenkomen. Er zijn echter ook zogenaamde lensparameters die de twee gebeurtenissen met elkaar verbinden. Voor een gegeven lensmodel hebben deze parameters verwachte verdelingen. Bijvoorbeeld, voor een sterrenstelsel zou de tijdvertraging tussen de beelden niet meer dan een paar maanden mogen zijn, met de grootste waarschijnlijkheid die tussen één en een paar dagen ligt. De kans dat niet-gelensde paren overeenkomstige intrinsieke parameters en lensparameters hebben is veel kleiner dan de kans dat alleen hun intrinsieke parameters overeenkomen. Daarom is het belangrijk om deze informatie in de zoekmethode toe te voegen. Over het algemeen zijn zoekalgoritmes onafhankelijk van lensmodellen om te voorkomen dat een gelensde gebeurtenis gemist wordt omdat die niet aan het gebruikte model voldoet. In dit proefschrift tonen we hoe men lensmodellen in het eindresultaat kan toevoegen zonder de hele zoekmethode opnieuw te moeten uitvoeren. Daardoor voorkomen we de intensieve procedure waarbij de analyse voor elk model herhaald moet worden. Bovendien tonen we een consequente vermindering van de kans op vals alarm als het juiste lensprofiel gebruikt wordt. We tonen ook dat kleine systematische afwijkingen van het lensmodeling de vermindering niet aanzienlijk tegenwerkt. In tegendeel, het gebruik van een verkeerd lensprofiel zou lensdetecties bijna onmogelijk maken. Op basis

van ons resultaten stellen we een zoekstrategie voor sterke lensing voor: we beginnen met de gebruikelijke modelonafhankelijke zoekmethode om de meest belovende kandidaten te identificeren. Vervolgens nemen we een reeks modellen op in de detectiestatistiek. Ze worden gekozen met redundantie tussen lensprofielen om rekening te houden met systematische effecten. Een echt lenspaar zou een hoge detectiestatistiek behouden voor alle lensmodellen van eenzelfde familie. Hoewel deze methode de detectie niet garandeert, is het een veilige manier om naar sterk gelensde paren te zoeken.

Vroegtijdige waarschuwingssystemen voor binaire neutronensterren met machine learning

In de afgelopen eeuw was gravitatiegolffysica niet het enige zoekgebied met doorbraken. Dat geldt ook voor machine learning, dat van een concept tot een krachtig algemeen instrument uitgegroeid is. Tegenwoordig is het rijp genoeg om in andere onderzoekgebieden toegepast te worden. Voor zwaartekrachtgolfanalyse is een van de meest wenselijke eigenschappen de snelheid na training. Men moet vaak computationeel dure methoden gebruiken vanwege de tijd die het algoritme nodig heeft om te convergeren of de noodzaak om vele parametercombinaties te doorlopen om een trigger te vinden. Met machine learning wordt de computationele kost stroomopwaarts verplaatst, met name tijdens de trainingfase. Eens het algoritme getraind is, kan het binnen enkele seconden resultaten produceren. Dit is wenselijk, vooral voor versnelde detecties. Deze zijn cruciaal voor dubbele neutronensterren, waar we een elektromagnetische emissie in verband met de vroegste momenten van de fusie willen observeren. De studie van het evenement in de beginfase is belangrijk om het gedrag van materie onder deze extreme omstandigheden beter te begrijpen.

In dit proefschrift bouwen we het eerste vroegtijdige waarschuwingsysteem voor binaire neutronensterren op basis van machine learning. We maken een convolutioneel neuraal netwerk dat een trigger zendt wanneer het een inspiraliserende binaire neutronenster opmerkt. In eerste instantie nemen we een verkleind tijdsvenster waarin het signaal op 20 Hz begint en splitsen we de binaire neutronenster in drie massacategorieën. We trainen en testen het netwerk op ontwerpsensitiviteit voor één LIGO-detector en stellen vast dat we het vroege deel van het signaal kunnen waarnemen, mits het luid genoeg is. In deze proof-of-concept studie is de opstelling niet echt realistisch omdat we de hemellocatie niet veranderen en het volledige detectornetwerk niet gebruiken. Daarom verbeteren we ons werklijn in een tweede stap. We verlagen de minimumfrequentie die door ons netwerk beschouwd wordt tot ongeveer 12 Hz. Bovendien vergroten we de vensterbreedte en laten we het beginfrequentie van het signaal variëren. Hierdoor zijn de massacategorieën niet meer nodig. We gebruiken ook een nieuwe trainingstechniek: curriculum learning, waarbij we de gegevens die het netwerk beschouwd tijden de trainingsfase iteratief moeilijker wordt. Nadat al deze verbeteringen zijn toegepast trainen en testen we het netwerk op verschillende detectorgeluiden: gesimuleerde ruis die

overeenstemt met de ruis voor de derde and vierde observatierondes en echte ruis van de derde observatieronde. We zien dat onze prestaties licht achteruitgaan voor de echte gegevens die zich minder goed gedragen. Toch behouden we enige detecteerbaarheid. Bovendien zorgt de verbeterde trainingsstrategie voor een hogere gevoeligheid. Voor een O4-achtig scenario zouden er ongeveer 20 binaire neutronensterrenfusies over één jaar gedetecteerd worden. Daarvan zou ons netwerk er ongeveer drie op voorhand detecteren, met in sommige gevallen een trigger 90 s voor de fusie. Helaas heeft ons algoritme ook een hoog percentage aan valse alarmen. Dit kan verminderd worden door meerdere triggers na elkaar te vragen. Deze oplossing verkort wel de tijd tussen de trigger en de samenvoeging. Kortom, dit werk toont aan dat machine learning een realistisch alternatief is voor vroegtijdige waarschuwing voor zwaartekrachtgolven, ook al zijn er extra ontwikkelingen nodig voordat het systeem praktisch bruikbaar is.

Naast een trigger die zegt dat er een fusie gaat gebeuren moeten astronomen ook weten waarnaar ze hun telescopen moeten richten om een elektromagnetische tegenhanger te zoeken. Daarom is het belangrijk om een snelle en precieze manier te hebben om een hemelkaart te maken. Dit is ook een plek waar machine learning gebruikt kan worden. Daarbij maakt men zich ook vaak zorgen over de betrouwbaarheid en de afhankelijk van de trainingsuis van machine learning. Daarom tonen we in dit proefschrift een hybride aanpak waar we machine learning en Bayesiaanse statistiek combineren om snel de hemelkaart en massadistributie van een evenement te verkrijgen. We gebruiken een neurale netwerk om goede voorstelverdelingen te verkrijgen voor de massa's en de hemellocatie. Vervolgens gebruiken we importance sampling om de verdelingen om te zetten naar de finale verdelingen. Uiteindelijk kan ons algoritme een hemelkaart produceren die vergelijkbaar is met die van traditionele Bayesiaanse methoden. Wij hebben echter slechts enkele minuten nodig vergeleken met dagen voor de traditionele methoden. Bovendien kan het aantal effectieve steekproeven gebruikt worden om de kwaliteit van de voorstelverdeling en het gedrag van het neurale netwerk te beoordelen. Dit systeem moet echter nog verbeterd worden om in de praktijk te werken. Onze voorgestelde voorstelverdeling voor de hemelkaart moet bijvoorbeeld ook rekening houden met hemelreflecties. In ieder geval, toont dit ook aan dat machine learning een veelbelovende weg is voor snelle localisatie en parameterschatting.

Als we het kader voor de bepaling van de hemelpositie en het vroegtijdige waarschuwingssysteem aanpassen om samen in realistische scenario's te werken, dan hebben we het eerste vroegtijdige waarschuwingssysteem voor binaire neutronensterren dat puur op basis van machine learning gebouwd is.

Derde generatie detectoren

Zelfs voordat de eerste zwaartekrachtgolf gedetecteerd was, was de planning voor de volgende generatie detectoren al begonnen. Simpel gesteld zijn deze detectoren groter en gevoeliger dan de huidige. Dit leidt tot meer observaties van compacte binaire fusie

signalen en de detectie van nieuwe fenomenen. De volgende generatie detectoren zijn de Europese Einstein Telescope, die meestal als een driehoekig netwerk met 10 km lange armen wordt aangezien, en de Amerikaanse Cosmic Explorer, die het L-vormige ontwerp behoudt, maar de armlengtes drastisch vergroot, tot 20 of 40 km. Naast hun grotere afmetingen zullen deze detectoren ook nieuwe technologieën gebruiken. Daardoor wordt hun bereik voor zwart gat-paren vergroot tot aan de kosmische horizon, waar alleen primordiale zwarte gaten – veroorzaakt door inhomogeniteiten kort na de Big Bang – zouden kunnen bestaan. Binaire neutronensterren zouden tot enkele roodverschuivingen gezien kunnen worden, waardoor meer multi-messenger detecties mogelijk zijn en een beter begrip van hun ruimtelijke verspreiding mogelijk wordt.

Door hun hogere gevoeligheid en grotere bereik zullen derde generatie interferometers meer gebeurtenissen detecteren, en signalen zullen langer zichtbaar zijn. Daarom vermoedt men dat ze in de detectieband elkaar zullen overlappen. In dit proefschrift bestuderen we eerst hoe vaak zulke overlappende signalen zullen voorkomen. Daarvoor simuleren we één jaar aan derde generatie detectornetwerk gegevens. Als we alleen naar binaire neutronensterren en zwarte gaten kijken, lijken overlappende signalen al relatief vaak voor te komen, met tientallen fusies die plaatsvinden terwijl een binaire neutronenster signaal in de detectieband is. Bovendien eindigen binnen een jaar wel tientallen tot duizenden fusies binnen dezelfde second, afhankelijk van de beschouwde lokale fusiesnelheid. Overlappende signalen zullen voorkomen wanneer de derde generatie detectoren in gebruik komen. Daarom bestuderen we hun invloed op de data-analyse. Daarvoor kijken we naar de overlapping van twee zwarte gaten paar signalen, twee binaire neutronenster signalen, en een paar zwarte gaten met een paar neutronenster signaal die tegelijkertijd of twee seconden na elkaar fuseren. Als we geen rekening houden met een van de twee signalen blijken er afwijkingen voor te komen, afhankelijk van de situatie. Afwijkingen komen voor in de geobserveerde parameters voor de binaire zwart gat signaal als het met een neutronenster signaal overlapt. Voor het zwart gat signaal met de hoogste totaal massa kunnen we het signaal helemaal niet waarnemen. Anderzijds zien we geen effect op de binaire neutronenster signaal. Voor het overlap van twee binaire neutronensterren signalen zien we dat we de parameters van het stilste signaal niet kunnen waarnemen. We observen geen afwijkingen voor het luidste signaal. In de andere gevallen zien we geen afwijkingen. Dit toont dat de huidige parameterschattingsmethoden in sommige gevallen werken maar in andere scenario's niet. Daarom moeten er aangepaste technieken ontwikkeld worden.

Kort na onze studie over afwijkingen in overlappende signalen hebben andere studies het effect bevestigd. In het bijzonder hebben overlappende binaire zwarte gaten met nabije kenmerken ook afwijkingen in de verdelingen indien hun fusietijden dicht bij elkaar liggen. Bovendien hangt de afwijking af van de precieze karakteristieken van de signalen. In dit proefschrift presenteren we twee mogelijke Bayesiaanse analysemethoden om overlappende signalen te analyseren: (i) hiërarchische subtractie, waar we het dominante signaal eerst analyseren, het resultaat aftrekken van de gegevens en daarna het tweede signaal bestuderen, en (ii) gezamenlijke parameterschatting, waarbij we de twee signalen

tegelijkertijd fitten. De tweede methode is nauwkeuriger omdat het op de correcte manier rekening houdt met de ruiskarakteristieken van de gegevens, maar de computationele kosten zijn hoger omdat de dimensie van de parameterruimte groter is. We vergelijken de resultaten van deze twee methoden met die verkregen wanneer de signalen zonder overlap geanalyseerd worden. Wegens computationele beperkingen concentreren we ons hier op binaire zwarte gaten met hoge massa's die 0.1 s uit elkaar fuseren. Zoals verwacht vinden we een afwijking bij de analyse van de dominante signaal door hiërarchische subtractie. Deze afwijking komt niet altijd terug in het tweede signaal. Daarom proberen we een derde stap toe te passen waarbij we de tweede signaal van de gegevens aftrekken en het eerste terug analyseren. Wij constateren geen significante verbetering in de herstelde waarschijnlijkheidsverdelingen. Voor gezamenlijke parameterschatting vinden we nauwkeurigere resultaten dan voor hiërarchische aftrekking. De resultaten zijn niet gelijkwaardig aan die van de afzonderlijke parameterschatting, waarschijnlijk door de toegenomen complexiteit in het probleem. Toch zou men met enkele verbeteringen nog betere resultaten met gezamenlijke parameterschatting kunnen bereiken. Deze studie is de allereerste die gezamenlijke analyse van twee zwaartekrachtgolfsignalen doet. Het maakt de weg vrij voor toekomstige ontwikkelingen.

Een van de nadelen van de Bayesiaanse methoden om overlappende signalen te analyseren is hun gebrek aan snelheid. Momenteel zouden ze de detectiesnelheid van de derde generatie detectoren niet kunnen volgen. Daarom hebben we in dit proefschrift ook een alternatieve analysemethode onderzocht. Het maakt gebruik van normalizing flows om overlappende binaire zwarte-gatsignalen gezamenlijk te analyseren. Met deze methode, die op concepten uit machine learning gebaseerd is, krijgen we de parameterverdelingen voor de zwarte gat-paren in een seconde na training, vergeleken met 20 tot 30 dagen voor gezamenlijke parameterschatting. We stellen dus vast dat ons algoritme de parameters van twee overlappende signalen correct kan schatten. Het kan dit consequent doen, ook al zien we meer variatie dan voor de parameterschatting van individuele signalen met normalizing flows. In vergelijking met de traditionele Bayesiaanse methoden heeft ons netwerk typisch bredere verdelingen. Vermoedelijk is de oorzaak de beperkte grootte van ons netwerk, dat tweemaal zo klein is als de netwerken gebruikt voor de parameterschatting van individuele signalen voor een dubbel zo grote parameterruimte. In plaats van het netwerk te vergroten – wat zou betekenen dat het niet meer op low-end GPUs kan werken – zouden we nauwkeurigere verdelingen kunnen krijgen door importance sampling op de output van het netwerk te gebruiken. Dit is echter moeilijk omdat we goed gecorreleerde gegevenspunten nodig hebben. Toch toont deze studie aan dat machine learning een haalbaar idee is om overlappende signalen te bestuderen, zelfs wanneer er rekening gehouden wordt met de hogere detectiesnelheid van de detectoren. Er zijn nog enkele verbeteringen nodig, maar deze techniek zou het mogelijk maken om studies voor derde generatie detectoren uit te voeren zonder op toekomstige ontwikkelingen van krachtigere methoden te moeten wachten.

Uiteindelijk zijn derde generatie detectoren instrumenten die onze kennis aanzienlijk vooruit gaan helpen. Er bestaan nog enkele knelpunten op het gebied van data-analyse.

Die moeten aangepakt worden om het volledige potentieel van deze interferometers te benutten. In dit proefschrift hebben we enkele van deze punten bestudeerd en hebben we mogelijke oplossingen voorgesteld.

CURRICULUM VITAE

Personal Information

Name: Justin Janquart
Nationality: Belgian
Birth Data: 12th of July 1997
Email: j.janquart@nikhef.nl / j.janquart@uu.nl / justin_janquart@hotmail.be

Education

2020 - 2023: Ph.D. in Physics
Nikhef & GRASP, Utrecht University; Prof. dr. C. Van Den Broeck

2018 - 2020: M.Phil in Space Science (summa cum laude)
Master Thesis: *Gravitational waves signal analysis: Matched filtering, typical analyses and beyond*⁵.
University of Liège; Prof. dr. J-R. Cudell

2015 - 2108: B.Sc. in Physics (magna cum laude)
University of Liège

Selected Publications

- J. Langendorff, A. Kolmus, J. Janquart, et al. “Normalizing flows as an avenue to study overlapping gravitational wave signals.” *arXiv e-prints*, arXiv:2211.15097, November 2022.
- J. Janquart, T. Baka, A. Samajdar, et al. “Parameter estimation methods for analyzing overlapping gravitational wave signals in the third-generation detector era.”, *arXiv e-prints*, arXiv:2211.01304, November 2022.
- J. Janquart, A. More, and C. Van Den Broeck. “Ordering the confusion: a study of the impact of lens models on gravitational-wave strong lensing detection capabilities.” *Mon. Not. Roy. Astron. Soc.*, 519(2):2046–2059, December 2022.
- A. Kolmus, G. Baltus, J. Janquart, et al. “Fast sky localization of gravitational waves using deep learning seeded importance sampling.” *Phys. Rev. D*, 106(2):023032, July 2022.

⁵<https://matheo.uliege.be/handle/2268.2/9211>

- G. Baltus, J. Janquart, M. Lopez, et al. “Convolutional neural network for gravitational-wave early alert: Going down in frequency.” *Phys. Rev. D*, 106(4):042002, May 2022.
- P. K. Gupta, A. Puecher, P. T. H. Pang, J. Janquart et al. “Determining the equation of state of neutron stars with Einstein Telescope using tidal effects and r-mode excitations from a population of binary inspirals.” *arXiv e-prints*, arXiv:2205.01182, May 2022.
- J. Janquart, E. Seo, O. A. Hannuksela, et al. “On the Identification of Individual Gravitational-wave Image Types of a Lensed System Using Higher-order Modes.” *Astrophys. J. Lett.*, 923(1):L1, December 2021.
- J. Janquart, O. A. Hannuksela, Haris K., et al. “A fast and precise methodology to search for and analyse strongly lensed gravitational-wave events.” *Mon. Not. Roy. Astron. Soc.*, 506(4):5430–5438, July 2021.
- A. Samajdar, J. Janquart, C. Van Den Broeck, et al. “Biases in parameter estimation from overlapping gravitational-wave signals in the third-generation detector era.” *Phys. Rev. D*, 104(4):044003, August 2021.
- G. Baltus, J. Janquart, M. Lopez, et al. “Convolutional neural networks for the detection of the early inspiral of a gravitational-wave signal.” *Phys. Rev. D*, 103:102003, May 2021

BIBLIOGRAPHY

- [1] A. Einstein. *Zur Elektrodynamik bewegter Körper*, pages 26–50. Vieweg+Teubner Verlag, Wiesbaden, December 1923.
- [2] Albert Einstein. Die Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften*, pages 844–847, January 1915.
- [3] R. V. Pound and G. A. Rebka. Gravitational redshift in nuclear resonance. *Phys. Rev. Lett.*, 3:439–441, November 1959.
- [4] F. W. Dyson, A. S. Eddington, and C. Davidson. A Determination of the Deflection of Light by the Sun’s Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919. *Phil. Trans. Roy. Soc. Lond. A*, 220:291–333, November 1920.
- [5] D. P. Bennett. Detection of Extrasolar Planets by Gravitational Microlensing. In John W. Mason, editor, *Exoplanets*, page 47. February 2008.
- [6] A. Gould. Recent Developments in Gravitational Microlensing. *ASP Conf. Ser.*, 403:86, March 2009.
- [7] M. W. Auger, T. Treu, A. S. Bolton, et al. The Sloan Lens ACS Survey. X. Stellar, Dynamical, and Total Mass Correlations of Massive Early-type Galaxies. *Astrophys. J.*, 724(1):511–525, November 2010.
- [8] M. Barnabè, O. Czoske, L. V. E. Koopmans, et al. Two-dimensional kinematics of SLACS lenses – III. Mass structure and dynamics of early-type lens galaxies beyond $z \sim 0.1$. *Mon. Not. Roy. Astron. Soc.*, 415(3):2215–2232, August 2011.
- [9] S. Vegetti, L. V. E. Koopmans, M. W. Auger, et al. Inference of the cold dark matter substructure mass function at $z = 0.2$ using strong gravitational lenses. *Mon. Not. Roy. Astron. Soc.*, 442(3):2017–2035, June 2014.
- [10] C. Grillo, M. Lombardi, and G. Bertin. Cosmological parameters from strong gravitational lensing and stellar dynamics in elliptical galaxies. *Astron. Astrophys.*, 477:397, January 2008.
- [11] M. Oguri, N. Inada, M. A. Strauss, et al. The Sloan Digital Sky Survey Quasar Lens Search. VI. Constraints on Dark Energy and the Evolution of Massive Galaxies. *Astrophys. J.*, 143(5):120, May 2012.
- [12] T. E. Collett and M. W. Auger. Cosmological constraints from the double source plane lens SDSSJ0946+1006. *Mon. Not. Roy. Astron. Soc.*, 443(2):969–976, September 2014.
- [13] S. H. Suyu, T. Treu, S. Hilbert, et al. Cosmology from gravitational lens time delays and Planck data. *Astrophys. J. Lett.*, 788:L35, May 2014.
- [14] A. M. Quider, M. Pettini, A. E. Shapley, et al. The ultraviolet spectrum of the gravitationally lensed galaxy ‘the Cosmic Horseshoe’: a close-up of a star-forming galaxy at $z \sim 2$. *Mon. Not. Roy. Astron. Soc.*, 398(3):1263–1278, September 2009.
- [15] E. R. Newton, P. J. Marshall, T. Treu, et al. The Sloan Lens ACS Survey. XI. Beyond Hubble Resolution: Size, Luminosity, and Stellar Mass of Compact Lensed Galaxies at Intermediate Redshift. *Astrophys. J.*, 734(2):104, June 2011.

- [16] The Event Horizon Telescope Collaboration. First m87 event horizon telescope results. i. the shadow of the supermassive black hole. *Astrophys. J. Lett.*, 875(1):L1, April 2019.
- [17] The Event Horizon Telescope Collaboration. First m87 event horizon telescope results. ii. array and instrumentation. *Astrophys. J. Lett.*, 875(1):L2, April 2019.
- [18] The Event Horizon Telescope Collaboration. First m87 event horizon telescope results. iii. data processing and calibration. *Astrophys. J. Lett.*, 875(1):L3, April 2019.
- [19] The Event Horizon Telescope Collaboration. First m87 event horizon telescope results. iv. imaging the central supermassive black hole. *Astrophys. J. Lett.*, 875(1):L4, April 2019.
- [20] The Event Horizon Telescope Collaboration. First m87 event horizon telescope results. v. physical origin of the asymmetric ring. *Astrophys. J. Lett.*, 875(1):L5, April 2019.
- [21] A. Einstein. Nherungsweise Integration der Feldgleichungen der Gravitation. *Sitzungsberichte der Kniglich Preussischen Akademie der Wissenschaften*, pages 688–696, January 1916.
- [22] R. A. Hulse and J. H. Taylor. Discovery of a pulsar in a binary system. *Astrophys. J. Lett.*, 195:L51–L53, January 1975.
- [23] J. M. Weisberg and J. H. Taylor. Relativistic binary pulsar B1913+16: Thirty years of observations and analysis. *ASP Conf. Ser.*, 328:25, July 2005.
- [24] M. Maggiore. *Gravitational Waves. Vol. 1: Theory and Experiments*. Oxford Master Series in Physics. Oxford University Press, October 2007.
- [25] The LIGO Scientific Collaboration. Advanced LIGO. *arXiv e-prints*, 32:074001, November 2015.
- [26] VIRGO Collaboration, F. Acernese, et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. *Class. Quant. Grav.*, 32(2):024001, January 2015.
- [27] K. Somiya. Detector configuration of KAGRA: The Japanese cryogenic gravitational-wave detector. *Class. Quant. Grav.*, 29:124007, March 2012.
- [28] Y. Aso, Y. Michimura, K. Somiya, et al. Interferometer design of the KAGRA gravitational wave detector. *Phys. Rev. D*, 88(4):043007, August 2013.
- [29] KAGRA Collaboration, T. Akutsu, et al. Overview of KAGRA: Detector design and construction history. *arXiv e-prints*, page arXiv:2005.05574, May 2020.
- [30] M. Punturo, M. Abernathy, F. Acernese, et al. The einstein telescope: a third-generation gravitational wave observatory. *Class. Quant. Grav.*, 27(19):194002, September 2010.
- [31] S. Hild, M. Abernathy, F. Acernese, et al. Sensitivity Studies for Third-Generation Gravitational Wave Observatories. *Class. Quant. Grav.*, 28:094013, April 2011.
- [32] LIGO Scientific Collaboration. LIGO’s Interferometer. https://www.ligo.caltech.edu/system/media_files/binaries/239/original/Basic_michelson_with_FP_and_PR_labeled.jpg?1435864227. Online, accessed October 2022.
- [33] LIGO Scientific Collaboration and Virgo Collaboration. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116:061102, February 2016.

- [34] LIGO Scientific Collaboration and Virgo Collaboration. Tests of general relativity with GW150914. *Phys. Rev. Lett.*, 116(22):221101, May 2016. [Erratum: *Phys. Rev. Lett.* 121, 129902 (2018)].
- [35] LIGO Scientific Collaboration and Virgo Collaboration. GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs. *Phys. Rev. X*, 9(3):031040, September 2019.
- [36] LIGO Scientific Collaboration and Virgo Collaboration. GW170814: A Three-Detector Observation of Gravitational Waves from a Binary Black Hole Coalescence. *Phys. Rev. Lett.*, 119(14):141101, October 2017.
- [37] LIGO Scientific Collaboration and Virgo Collaboration. GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral. *Phys. Rev. Lett.*, 119(16):161101, October 2017.
- [38] LIGO Scientific Collaboration, Virgo Collaboration, Fermi GBM, et al. Multi-messenger Observations of a Binary Neutron Star Merger. *Astrophys. J. Lett.*, 848(2):L12, October 2017.
- [39] LIGO Scientific Collaboration and Virgo Collaboration. Tests of General Relativity with GW170817. *Phys. Rev. Lett.*, 123(1):011102, July 2019.
- [40] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run. November 2021.
- [41] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. GW190412: Observation of a Binary-Black-Hole Coalescence with Asymmetric Masses. *Phys. Rev. D*, 102(4):043015, August 2020.
- [42] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. GW190425: Observation of a Compact Binary Coalescence with Total Mass $\sim 3.4M_{\odot}$. *Astrophys. J. Lett.*, 892(1):L3, March 2020.
- [43] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. GW190521: A Binary Black Hole Merger with a Total Mass of $150M_{\odot}$. *Phys. Rev. Lett.*, 125(10):101102, September 2020.
- [44] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. GW190814: Gravitational Waves from the Coalescence of a 23 Solar Mass Black Hole with a 2.6 Solar Mass Compact Object. *Astrophys. J. Lett.*, 896(2):L44, June 2020.
- [45] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. Observation of Gravitational Waves from Two Neutron Star–Black Hole Coalescences. *Astrophys. J. Lett.*, 915(1):L5, July 2021.
- [46] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. The population of merging compact binaries inferred using gravitational waves through GWTC-3. *arXiv e-prints*, page 2111.03634, November 2021.
- [47] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. Tests of General Relativity with GWTC-3. *arXiv e-prints*, page 2112.06861, December 2021.
- [48] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. Constraints on the cosmic expansion history from GWTC-3. *arXiv e-prints*, page 2111.03604, November 2021.

[49] B. Iyer, T. Souradeep, C. Unnikrishnan, et al. LIGO-India, Proposal of the Consortium for Indian Initiative in Gravitational-wave Observations (IndIGO). <https://dcc.ligo.org/LIGO-M1100296/public>. Online, accessed October 2022.

[50] D. Reitze, R. Adhikari, S. Ballmer, et al. Cosmic Explorer: The U.S. Contribution to Gravitational-Wave Astronomy beyond LIGO. *Bull. Am. Astron. Soc.*, 51(7):035, September 2019.

[51] M. Evans, R. Adhikari, A. Chaitanya, et al. A Horizon Study for Cosmic Explorer: Science, Observatories, and Community. *arXiv e-prints*, page 2109.08882, September 2021.

[52] M. Maggiore, C. Van Den Broeck, N. Bartolo, et al. Science Case for the Einstein Telescope. *JCAP*, 03:050, March 2020.

[53] S. Babak, A. Petiteau, and M. Hewitson. LISA Sensitivity and SNR Calculations. *arXiv e-prints*, page 2108.01167, August 2021.

[54] G. Hobbs and S. Dai. Gravitational wave research using pulsar timing arrays. *Natl. Sci. Rev.*, 4(5):707–717, December 2017.

[55] Md F. Alam, Z. Arzoumanian, P.T. Baker, et al. The NANOGrav 12.5 yr Data Set: Wideband Timing of 47 Millisecond Pulsars. *Astrophys. J. Suppl.*, 252(1):5, January 2021.

[56] Z. Arzoumanian, P.T. Baker, A. Brazier, et al. Multimessenger Gravitational-wave Searches with Pulsar Timing Arrays: Application to 3C 66B Using the NANOGrav 11-year Data Set. *Astrophys. J.*, 900(2):102, September 2020.

[57] N. S. Pol, S.R. Taylor, L. Zoltan Kelley, et al. Astrophysics Milestones for Pulsar Timing Array Gravitational-wave Detection. *Astrophys. J. Lett.*, 911(2):L34, April 2021.

[58] Z. Arzoumanian, T.P Baker, B. Harsha, et al. The NANOGrav 12.5 yr Data Set: Search for an Isotropic Stochastic Gravitational-wave Background. *Astrophys. J. Lett.*, 905(2):L34, December 2020.

[59] Z. Arzoumanian, P.T Baker, B. Harsha, et al. Searching for Gravitational Waves from Cosmological Phase Transitions with the NANOGrav 12.5-Year Dataset. *Phys. Rev. Lett.*, 127(25):251302, December 2021.

[60] Z. Arzoumanian, P.T Baker, B. Harsha, et al. The NANOGrav 12.5-year Data Set: Search for Non-Einsteinian Polarization Modes in the Gravitational-wave Background. *Astrophys. J. Lett.*, 923(2):L22, December 2021.

[61] C. Moore, R. Cole, C. Berry, et al. GWplotter: Gravitational Wave Detectors and Sources. <http://gwplotter.com/>. Online, accessed November 2022.

[62] E. Cuoco, J. Powell, M. Cavaglià, et al. Enhancing Gravitational-Wave Science with Machine Learning. *Mach. Learn. Sci. Tech.*, 2(1):011002, May 2021.

[63] A. Le Tiec and J. Novak. *An Overview of Gravitational Waves: Theory, Sources and Detection*, page 1607.04202. July 2017.

[64] B. F. Schutz. Networks of gravitational wave detectors and three figures of merit. *Class. Quant. Grav.*, 28:125023, June 2011.

[65] Extracting Physics from Gravitational Waves: Testing the Strong-field Dynamics of General Relativity and Inferring the Large-scale Structure of the Universe. https://www.nikhef.nl/pub/services/biblio/theses_pdf/thesis_T_G_F_Li.pdf. Online, accessed October 2022.

[66] N. Sarin and P. D. Lasky. The evolution of binary neutron star post-merger remnants: a review. *Gen. Rel. Grav.*, 53(6):59, June 2021.

[67] T. Damour, B. R. Iyer, and B. S. Sathyaprakash. A Comparison of search templates for gravitational waves from binary inspiral. *Phys. Rev. D*, 63:044023, January 2001. [Erratum: Phys. Rev. D 72, 029902 (2005)].

[68] Thibault Damour, Bala R. Iyer, and B. S. Sathyaprakash. A Comparison of search templates for gravitational waves from binary inspiral - 3.5PN update. *Phys. Rev. D*, 66:027502, July 2002.

[69] T. Futamase and Y. Itoh. The post-newtonian approximation for relativistic compact binaries. *Living Reviews in Relativity*, 10, March 2007.

[70] A. Buonanno, B. Iyer, E. Ochsner, et al. Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors. *Phys. Rev. D*, 80:084043, October 2009.

[71] L. E. Kidder, C. M. Will, and A. G. Wiseman. Spin effects in the inspiral of coalescing compact binaries. *Phys. Rev. D*, 47(10):R4183–R4187, May 1993.

[72] B. S. Sathyaprakash and S. V. Dhurandhar. Choice of filters for the detection of gravitational waves from coalescing binaries. *Phys. Rev. D*, 44:3819–3834, December 1991.

[73] A. Lundgren and R. O’Shaughnessy. Single-spin precessing gravitational waveform in closed form. *Phys. Rev. D*, 89(4):044021, February 2014.

[74] F. Pretorius. Evolution of binary black-hole spacetimes. *Phys. Rev. Lett.*, 95:121101, September 2005.

[75] B. Bruegmann, J. A. Gonzalez, M. Hannam, et al. Calibration of Moving Puncture Simulations. *Phys. Rev. D*, 77:024027, January 2008.

[76] M. Thierfelder, S. Bernuzzi, and B. Bruegmann. Numerical relativity simulations of binary neutron stars. *Phys. Rev. D*, 84:044012, August 2011.

[77] D. Radice and L. Rezzolla. Thc: A new high-order finite-difference high-resolution shock-capturing code for special-relativistic hydrodynamics. *Astron. Astrophys.*, 547, October 2012.

[78] L. Kidder, H. Pfeiffer, and M. Scheek. Spectral Einstein Code. <https://www.black-holes.org/code/SpEC.html>.

[79] L. Baiotti, I. Hawke, P. J. Montero, et al. Three-dimensional relativistic simulations of rotating neutron star collapse to a Kerr black hole. *Phys. Rev. D*, 71:024035, January 2005.

[80] Y. Zlochower, J. G. Baker, Manuela Campanelli, et al. Accurate black hole evolutions by fourth-order numerical relativity. *Phys. Rev. D*, 72:024021, July 2005.

[81] T. Yamamoto, M. Shibata, and K. Taniguchi. Simulating coalescing compact binaries by a new code SACRA. *Phys. Rev. D*, 78:064054, September 2008.

[82] A. H. Mroue, M.A. Scheel, B. Szilagyi, et al. Catalog of 174 Binary Black Hole Simulations for Gravitational Wave Astronomy. *Phys. Rev. Lett.*, 111(24):241104, December 2013.

[83] M. Boyle, D.A.B Hemberger, G. Lovelace, et al. The SXS Collaboration catalog of binary black hole simulations. *Class. Quant. Grav.*, 36(19):195006, October 2019.

[84] K. Jani, J. Healy, J. A. Clark, et al. Georgia Tech Catalog of Gravitational Waveforms. *Class. Quant. Grav.*, 33(20):204001, October 2016.

[85] J. Healy, C. O. Lousto, J. Lange, et al. Second RIT binary black hole simulations catalog and its application to gravitational waves parameter estimation. *Phys. Rev. D*, 100(2):024021, July 2019.

[86] P. Ajith, M. Boyle, D. A. Brown, et al. The ninja-2 catalog of hybrid post-newtonian/numerical-relativity waveforms for non-precessing black-hole binaries. *Class. Quant. Grav.*, 29(12):124001, June 2012.

[87] T. Dietrich, D. Radice, S. Bernuzzi, et al. CoRe database of binary neutron star merger waveforms. *Class. Quant. Grav.*, 35(24):24LT01, December 2018.

[88] K. Kiuchi, K. Kawaguchi, K. Kyutoku, et al. Sub-radian-accuracy gravitational waves from coalescing binary neutron stars in numerical relativity. II. Systematic study on the equation of state, binary mass, and mass ratio. *Phys. Rev. D*, 101(8):084006, April 2020.

[89] D. Radice, L. Rezzolla, and F. Galeazzi. High-Order Numerical-Relativity Simulations of Binary Neutron Stars. *ASP Conf. Ser.*, 498:121–126, October 2015.

[90] M. Pürer. Frequency domain reduced order model of aligned-spin effective-one-body waveforms with generic mass-ratios and spins. *Phys. Rev. D*, 93(6):064041, March 2016.

[91] S. E. Field, C. R. Galley, J. S. Hesthaven, et al. Fast prediction and evaluation of gravitational waveforms using surrogate models. *Phys. Rev. X*, 4(3):031006, July 2014.

[92] J. Blackman, S. E. Field, C. R. Galley, et al. Fast and Accurate Prediction of Numerical Relativity Waveforms from Binary Black Hole Coalescences Using Surrogate Models. *Phys. Rev. Lett.*, 115(12):121102, September 2015.

[93] J. Blackman, S. E. Field, M. A. Scheel, et al. A Surrogate Model of Gravitational Waveforms from Numerical Relativity Simulations of Precessing Binary Black Hole Mergers. *Phys. Rev. D*, 95(10):104023, May 2017.

[94] J. Blackman, S. E. Field, M. A. Scheel, et al. Numerical relativity waveform surrogate model for generically precessing binary black hole mergers. *Phys. Rev. D*, 96(2):024058, July 2017.

[95] A. Buonanno and T. Damour. Effective one-body approach to general relativistic two-body dynamics. *Phys. Rev. D*, 59:084006, April 1999.

[96] A. Buonanno and T. Damour. Transition from inspiral to plunge in binary black hole coalescences. *Phys. Rev. D*, 62:064015, August 2000.

[97] T. Damour, P. Jaranowski, and G. Schaefer. On the determination of the last stable orbit for circular general relativistic binaries at the third post-Newtonian approximation. *Phys. Rev. D*, 62:084011, September 2000.

[98] A. Buonanno, Y. Pan, J. G. Baker, et al. Toward faithful templates for non-spinning binary black holes using the effective-one-body approach. *Phys. Rev. D*, 76:104049, November 2007.

[99] A. Bohé, L. Shao, A. Taracchini, et al. Improved effective-one-body model of spinning, nonprecessing binary black holes for the era of gravitational-wave astrophysics with advanced detectors. *Phys. Rev. D*, 95(4):044028, February 2017.

[100] R. Gamba, S. Akçay, S. Bernuzzi, et al. Effective-one-body waveforms for precessing coalescing compact binaries with post-Newtonian twist. *Phys. Rev. D*, 106(2):024020, July 2022.

[101] S. Akçay, S. Bernuzzi, F. Messina, et al. Effective-one-body multipolar waveform for tidally interacting binary neutron stars up to merger. *Phys. Rev. D*, 99(4):044051, February 2019.

[102] B. D. Lackey, M. Pürller, A. Taracchini, et al. Surrogate model for an aligned-spin effective one body waveform model of binary neutron star inspirals using Gaussian process regression. *Phys. Rev. D*, 100(2):024002, July 2019.

[103] S. Husa, S. Khan, M. Hannam, et al. Frequency-domain gravitational waves from nonprecessing black-hole binaries. I. New numerical waveforms and anatomy of the signal. *Phys. Rev. D*, 93(4):044006, February 2016.

[104] S. Khan, S. Husa, M. Hannam, et al. Frequency-domain gravitational waves from nonprecessing black-hole binaries. II. A phenomenological model for the advanced detector era. *Phys. Rev. D*, 93(4):044007, February 2016.

[105] K. G. Arun, A. Buonanno, G. Faye, et al. Higher-order spin effects in the amplitude and phase of gravitational waveforms emitted by inspiraling compact binaries: Ready-to-use gravitational waveforms. *Phys. Rev. D*, 79:104023, May 2009. [Erratum: *Phys. Rev. D* 84, 049901 (2011)].

[106] P. Schmidt, M. Hannam, and S. Husa. Towards models of gravitational waveforms from generic binaries: A simple approximate mapping between precessing and non-precessing inspiral signals. *Phys. Rev. D*, 86:104063, November 2012.

[107] M. Hannam, P. Schmidt, A. Bohé, et al. Simple model of complete precessing black-hole-binary gravitational waveforms. *Phys. Rev. Lett.*, 113(15):151101, October 2014.

[108] E.P. Wigner and H.S.W. Massey. *Group Theory: And Its Application to the Quantum Mechanics of Atomic Spectra*. Elsevier Science, January 2013.

[109] J. N. Goldberg, A. J. MacFarlane, E. T. Newman, et al. Spin s spherical harmonics and edth. *J. Math. Phys.*, 8:2155, September 1967.

[110] T. Dietrich, S. Bernuzzi, and W. Tichy. Closed-form tidal approximants for binary neutron star gravitational waveforms constructed from high-resolution numerical relativity simulations. *Phys. Rev. D*, 96(12):121501, December 2017.

[111] T. Dietrich, S. Khan, R. Dudi, et al. Matter imprints in waveform models for neutron star binaries: Tidal and self-spin effects. *Phys. Rev. D*, 99(2):024029, January 2019.

[112] T. Islam, A. K. Mehta, A. Ghosh, et al. Testing the no-hair nature of binary black holes using the consistency of multipolar gravitational radiation. *Phys. Rev. D*, 101(2):024032, January 2020.

[113] T. Islam. Applying higher-modes consistency test on GW190814 : lessons on no-hair theorem, nature of the secondary compact object and waveform modeling. October 2021.

[114] A. Puecher, C. Kalaghatgi, S. Roy, et al. Testing general relativity using higher-order modes of gravitational waves from binary black holes. *Phys. Rev. D*, 106(8):082003, October 2022.

[115] Y. Wang, R. K. L. Lo, A. K. Y. Li, et al. Identifying Type II Strongly Lensed Gravitational-Wave Images in Third-Generation Gravitational-Wave Detectors. *Phys. Rev. D*, 103(10):104055, May 2021.

[116] A. Vijaykumar, A. K. Mehta, and A. Ganguly. Detection and parameter estimation challenges of Type-II lensed binary black hole signals. *arXiv e-prints*, page 2202.06334, February 2022.

[117] J. Januart, E. Seo, O. A. Hannuksela, et al. On the Identification of Individual Gravitational-wave Image Types of a Lensed System Using Higher-order Modes. *Astrophys. J. Lett.*, 923(1):L1, December 2021.

[118] C.K. Mishra, A. Kela, K. G. Arun, et al. Ready-to-use post-Newtonian gravitational waveforms for binary black holes with non-precessing spins: An update. *Phys. Rev. D*, 93(8):084054, April 2016.

[119] S. Ossokine, A. Buonanno, S. Marsat, et al. Multipolar Effective-One-Body Waveforms for Precessing Binary Black Holes: Construction and Validation. *Phys. Rev. D*, 102(4):044055, August 2020.

[120] L. London, S. Khan, E. Fauchon-Jones, et al. First higher-multipole model of gravitational waves from spinning and coalescing black-hole binaries. *Phys. Rev. Lett.*, 120(16):161102, April 2018.

[121] G. Pratten, C. García-Quirós, M. Colleoni, et al. Computationally efficient models for the dominant and subdominant harmonic modes of precessing binary black holes. *Phys. Rev. D*, 103(10):104056, May 2021.

[122] D.V. Martynov, E.D. Hall, B.P Abbott, et al. Sensitivity of the Advanced LIGO detectors at the beginning of gravitational wave astronomy. *Phys. Rev. D*, 93(11):112004, June 2016. [Addendum: *Phys. Rev. D* 97, 059901 (2018)].

[123] L. Blackburn, L. Cadonati, S. Caride, et al. The LSC Glitch Group: Monitoring Noise Transients during the fifth LIGO Science Run. *Class. Quant. Grav.*, 25:184004, September 2008.

[124] LIGO Scientific Collaboration and Virgo Collaboration. Characterization of transient noise in Advanced LIGO relevant to gravitational wave signal GW150914. *Class. Quant. Grav.*, 33(13):134001, June 2016.

[125] M. Cabero, A. Lundgren, A.H. Nitz, et al. Blip glitches in Advanced LIGO data. *Class. Quant. Grav.*, 36(15):15, July 2019.

[126] D. Davis, J.S. Areeda, B.K. Berger, et al. LIGO detector characterization in the second and third observing runs. *Class. Quant. Grav.*, 38(13):135014, June 2021.

[127] N. Wiener. Generalized harmonic analysis. *Acta Mathematica*, 55:117–258, December 1930.

[128] B. S. Sathyaprakash and B. F. Schutz. Physics, Astrophysics and Cosmology with Gravitational Waves. *Living Rev. Rel.*, 12:2, December 2009.

[129] D. Mukherjee, S. Caudill, R. Magee, et al. Template bank for spinning compact binary mergers in the second observation run of Advanced LIGO and the first observation run of Advanced Virgo. *Phys. Rev. D.*, 103(8):084047, April 2021.

[130] K. Cannon, S. Caudill, C. Chan, et al. GstLAL: A software framework for gravitational wave discovery. *SoftwareX*, 14:100680, June 2021.

- [131] LIGO Scientific Collaboration, Virgo Collaboration, and Kagra Collaboration. GstLAL software. <https://git.ligo.org/lscsoft/gstlal>. Online, accessed November 2022.
- [132] B. Allen, W. G. Anderson, P. R. Brady, et al. FINDCHIRP: An Algorithm for detection of gravitational waves from inspiraling compact binaries. *Phys. Rev. D*, 85:122006, June 2012.
- [133] B. Allen. χ^2 time-frequency discriminator for gravitational wave detection. *Phys. Rev. D*, 71:062001, 2005.
- [134] T. Dal Canton, A.H. Nitz, A.P. Lundgren, et al. Implementing a search for aligned-spin neutron star-black hole systems with advanced ground based gravitational wave detectors. *Phys. Rev. D*, 90(8):082004, October 2014.
- [135] S. A. Usman, A.H. Nitz, I.W. Harry, et al. The PyCBC search for gravitational waves from compact binary coalescence. *Class. Quant. Grav.*, 33(21):215004, October 2016.
- [136] A. H. Nitz, T. Dent, T. Dal Canton, et al. Detecting binary compact-object mergers with gravitational waves: Understanding and Improving the sensitivity of the PyCBC search. *Astrophys. J.*, 849(2):118, November 2017.
- [137] G. S. Davies, T. Dent, M. Tápai, et al. Extending the PyCBC search for gravitational waves from compact binary mergers to a global network. *Phys. Rev. D*, 102(2):022004, July 2020.
- [138] T. Adams, D. Buskulic, V. Germain, et al. Low-latency analysis pipeline for compact binary coalescences in the advanced gravitational wave detector era. *Class. Quant. Grav.*, 33(17):175012, September 2016.
- [139] F. Aubin, F. Brightenti, R. Chierici, et al. The MBTA pipeline for detecting compact binary coalescences in the third LIGO–Virgo observing run. *Class. Quant. Grav.*, 38(9):095004, April 2021.
- [140] J. Luan, S. Hooper, L. Wen, et al. Towards low-latency real-time detection of gravitational waves from compact binary coalescences in the era of advanced detectors. *Phys. Rev. D*, 85:102002, May 2012.
- [141] Qi Chu, M. Kovalam, L. Wen, et al. SPIIR online coherent pipeline to search for gravitational waves from compact binary coalescences. *Phys. Rev. D*, 105(2):024023, January 2022.
- [142] S. Klimenko and G. Mitselmakher. A wavelet method for detection of gravitational wave bursts. *Class. Quant. Grav.*, 21(20):S1819, September 2004.
- [143] S. Klimenko, G. Vedovato, M. Drago, et al. Method for detection and reconstruction of gravitational wave transients with networks of advanced detectors. *Phys. Rev. D*, 93(4):042004, February 2016.
- [144] P. Godwin, R. Essick, C. Hanna, et al. Incorporation of Statistical Data Quality Information into the GstLAL Search Analysis. *arXiv e-prints*, page 2010.15282, October 2020.
- [145] W. M. Farr, J. R. Gair, I. Mandel, et al. Counting and confusion: Bayesian rate estimation with multiple populations. *Phys. Rev. D*, 91(2):023005, January 2015.
- [146] T. Dent. Technical note: Extending the PyCBC pastro calculation to a global network. <https://dcc.ligo.org/LIGO-T2100060/public>. Online, accessed November 2022.

[147] J. Skilling. Nested sampling for general Bayesian computation. *Bayesian Analysis*, 1(4):833 – 859, December 2006.

[148] J. Veitch, V. Raymond, B. Farr, et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Phys. Rev. D*, 91(4):042003, February 2015.

[149] M. Ashton, G. ad Huebner, P.D. Lasky, et al. BILBY: A user-friendly Bayesian inference library for gravitational-wave astronomy. *Astrophys. J. Suppl.*, 241(2):27, April 2019.

[150] I. M. Romero-Shaw, C. Talbot, S. Biscoveanu, et al. Bayesian inference for compact binary coalescences with BILBY: validation and application to the first LIGO–Virgo gravitational-wave transient catalogue. *Mon. Not. Roy. Astron. Soc.*, 499(3):3295–3319, December 2020.

[151] J. S. Speagle. DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. *Mon. Not. Roy. Astron. Soc.*, 493(3):3132–3158, April 2020.

[152] E. Higson, W. Handley, M. Hobson, et al. Dynamic nested sampling: an improved algorithm for parameter estimation and evidence calculation. *Statistics and Computing*, 29(5):891–913, September 2019.

[153] J. Buchner, A. Georgakakis, K. Nandra, et al. X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue. *Astron. Astrophys.*, 564:A125, April 2014.

[154] F. Feroz and M. P. Hobson. Multimodal nested sampling: an efficient and robust alternative to MCMC methods for astronomical data analysis. *Mon. Not. Roy. Astron. Soc.*, 384:449, February 2008.

[155] F. Feroz, J. R. Gair, M. P. Hobson, et al. Use of the MultiNest algorithm for gravitational wave data analysis. *Class. Quant. Grav.*, 26:215003, October 2009.

[156] W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, April 1970.

[157] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, et al. Equation of state calculations by fast computing machines. March 1953.

[158] J. Lee, W. Sung, and J. Choi. Metamodel for efficient estimation of capacity-fade uncertainty in li-ion batteries for electric vehicles. *Energies*, 8:5538–5554, June 2015.

[159] D. J. Earl and M. W. Deem. Parallel tempering: Theory, applications, and new perspectives. *Physical Chemistry Chemical Physics*, 7(23):3910, January 2005.

[160] T. B. Littenberg and N. J. Cornish. A Bayesian Approach to the Detection Problem in Gravitational Wave Astronomy. *Phys. Rev. D*, 80:063007, September 2009.

[161] G. Ashton and C. Talbot. Bilby-MCMC: an MCMC sampler for gravitational-wave inference. *Mon. Not. Roy. Astron. Soc.*, 507(2):2037–2051, October 2021.

[162] A. Einstein. A lens-like action of a star by the deviation of light in the gravitational field. *Science*, 84:506–507, December 1936.

[163] F. Zwicky. Nebulae as gravitational lenses. *Phys. Rev.*, 51:290–290, February 1937.

- [164] P. Schneider, C. Kochanek, and J. Wambsganss. *Gravitational Lensing: Strong, Weak and Micro*. Springer, January 2006.
- [165] H. C. Ohanian. On the Focusing of Gravitational Radiation. *International Journal of Theoretical Physics*, 9(6):425–437, June 1974.
- [166] S. Deguchi and W. D. Watson. Wave effects in gravitational lensing of electromagnetic radiation. *Phys. Rev. D*, 34:1708–1718, September 1986.
- [167] T. T. Nakamura. Gravitational lensing of gravitational waves from inspiraling binaries by a point mass lens. *Phys. Rev. Lett.*, 80:1138–1141, February 1998.
- [168] R. Takahashi and T. Nakamura. Wave effects in gravitational lensing of gravitational waves from chirping binaries. *Astrophys. J.*, 595:1039–1051, October 2003.
- [169] L. Dai and T. Venumadhav. On the waveforms of gravitationally lensed gravitational waves. *arXiv e-prints*, page 1702.04724, 2 2017.
- [170] J. M. Ezquiaga, D. E. Holz, W. Hu, et al. Phase effects from strong gravitational lensing of gravitational waves. *Phys. Rev. D*, 103(6):064047, March 2021.
- [171] K. Lai, O. A. Hannuksela, A. Herrera-Martín, et al. Discovering intermediate-mass black hole lenses through gravitational wave lensing. *Phys. Rev. D*, 98(8):083005, October 2018.
- [172] O. A. Hannuksela, K. Haris, K. K. Y. Ng, et al. Search for gravitational lensing signatures in LIGO-Virgo binary black hole events. *Astrophys. J. Lett.*, 874(1):L2, March 2019.
- [173] M. H. Y. Cheung, J. Gais, O. A. Hannuksela, et al. Stellar-mass microlensing of gravitational waves. *Mon. Not. Roy. Astron. Soc.*, 503(3):3326–3336, May 2021.
- [174] M. Wright and M. Hendry. Gravelamps: Gravitational Wave Lensing Mass Profile Model Selection. *Astrophys. J.*, 935(2):68, August 2022.
- [175] S. Jung and C.S. Shin. Gravitational-Wave Fringes at LIGO: Detecting Compact Dark Matter by Gravitational Lensing. *Phys. Rev. Lett.*, 122(4):041103, January 2019.
- [176] J. M. Diego. Constraining the abundance of primordial black holes with gravitational lensing of gravitational waves at LIGO frequencies. *Phys. Rev. D*, 101(12):123512, June 2020.
- [177] M. Oguri and R. Takahashi. Probing Dark Low-mass Halos and Primordial Black Holes with Frequency-dependent Gravitational Lensing Dispersions of Gravitational Waves. *Astrophys. J.*, 901(1):58, September 2020.
- [178] A. Mishra, A. K. Meena, A. More, et al. Gravitational lensing of gravitational waves: effect of microlens population in lensing galaxies. *Month. Not. Roy. Astron. Soc.*, 508(4):4869–4886, December 2021.
- [179] A. Liu, I. C. F. Wong, S. H. W. Leong, et al. Exploring the hidden Universe: A novel phenomenological approach for recovering arbitrary gravitational-wave millilensing configurations. February 2023.
- [180] K. K. Y. Ng, K. W. K. Wong, T. Broadhurst, et al. Precise LIGO Lensing Rate Predictions for Binary Black Holes. *Phys. Rev. D*, 97(2):023012, January 2018.

- [181] K. Haris, A. K. Mehta, S. Kumar, et al. Identifying strongly lensed gravitational wave signals from binary black hole mergers. *arXiv e-prints*, page 1807.07062, July 2018.
- [182] G. P. Smith, M. Jauzac, J. Veitch, et al. What if LIGO’s gravitational wave detections are strongly lensed by massive galaxy clusters? *Mon. Not. Roy. Astron. Soc.*, 475(3):3823–3828, April 2018.
- [183] G.P. Smith, C.P.L Berry, M. Bianconi, et al. Strong-lensing of Gravitational Waves by Galaxy Clusters. *IAU Symp.*, 338:98–102, October 2017.
- [184] G. P. Smith, A. Robertson, M. Bianconi, et al. Discovery of Strongly-lensed Gravitational Waves - Implications for the LSST Observing Strategy. *arXiv e-prints*, page arXiv:1902.05140, February 2019.
- [185] A. Robertson, G. P. Smith, R. Massey, et al. What does strong gravitational lensing? The mass and redshift distribution of high-magnification lenses. *Mon. Not. Roy. Astron. Soc.*, 495(4):3727–3739, July 2020.
- [186] D. Ryczanowski, G. P. Smith, M. Bianconi, et al. On building a cluster watch-list for identifying strongly lensed supernovae, gravitational waves and kilonovae. *Mon. Not. Roy. Astron. Soc.*, 495(2):1666–1671, June 2020.
- [187] R. A. C. Wierda, E. Wempe, O. A. Hannuksela, et al. Beyond the Detector Horizon: Forecasting Gravitational-Wave Strong Lensing. *Astrophys. J.*, 921(2):154, November 2021.
- [188] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. Search for Lensing Signatures in the Gravitational-Wave Observations from the First Half of LIGO–Virgo’s Third Observing Run. *Astrophys. J.*, 923(1):14, December 2021.
- [189] L. Dai, B. Zackay, T. Venumadhav, et al. Search for Lensed Gravitational Waves Including Morse Phase Information: An Intriguing Candidate in O2. *arXiv e-prints*, page 2007.12709, 7 2020.
- [190] E. Seo, O. A. Hannuksela, and T. G. F. Li. Improving Detection of Gravitational-wave Microlensing Using Repeated Signals Induced by Strong Lensing. *Astrophys. J.*, 932(1):50, June 2022.
- [191] J. Janquart, O. A. Hannuksela, Haris K., et al. A fast and precise methodology to search for and analyse strongly lensed gravitational-wave events. *Mon. Not. Roy. Astron. Soc.*, 506(4):5430–5438, July 2021.
- [192] O. A. Hannuksela, T. E. Collett, M. Çalışkan, et al. Localizing merging black holes with sub-arcsecond precision using gravitational-wave lensing. *Mon. Not. Roy. Astron. Soc.*, 498(3):3395–3402, August 2020.
- [193] E. Wempe, L. V. E. Koopmans, R. A. C. Wierda, et al. A lensing multi-messenger channel: Combining LIGO-Virgo-Kagra lensed gravitational-wave measurements with Euclid observations. *arXiv e-prints*, page 2204.08732, April 2022.
- [194] M. Sereno, P. Jetzer, A. Sesana, et al. Cosmography with strong lensing of LISA gravitational wave sources. *Mon. Not. Roy. Astron. Soc.*, 415:2773, August 2011.
- [195] K. Liao, X. Fan, X. Ding, et al. Precision cosmology from future lensed gravitational wave and electromagnetic signals. *Nature Commun.*, 8(1):1148, December 2017. [Erratum: *Nature Commun.* 8, 2136 (2017)].

- [196] S. Cao, J. Qi, Z. Cao, et al. Direct test of the FLRW metric from strongly lensed gravitational wave observations. *Sci. Rep.*, 9(1):11608, July 2019.
- [197] Y. Li, X. Fan, and L. Gou. Constraining Cosmological Parameters in the FLRW Metric with Lensed GW+EM Signals. *Astrophys. J.*, 873(1):37, March 2019.
- [198] T. Baker and M. Trodden. Multimessenger time delays from lensed gravitational waves. *Phys. Rev. D*, 95(6):063512, March 2017.
- [199] X. Fan, K. Liao, M. Biesiada, et al. Speed of Gravitational Waves from Strongly Lensed Gravitational Waves and Electromagnetic Signals. *Phys. Rev. Lett.*, 118(9):091102, March 2017.
- [200] S. Goyal, K. Haris, A.K. Mehta, et al. Testing the nature of gravitational-wave polarizations using strongly lensed signals. *Phys. Rev. D*, 103:024038, January 2021.
- [201] M. Çalışkan, J.M. Ezquiaga, O. A. Hannuksela, et al. Lensing or luck? False alarm probabilities for gravitational lensing of gravitational waves. *arXiv e-prints*, page 2201.04619, January 2022.
- [202] J. Januart, A. More, and C. Van Den Broeck. Ordering the confusion: a study of the impact of lens models on gravitational-wave strong lensing detection capabilities. *Mon. Not. Roy. Astro. Soc.*, 519(2):2046–2059, December 2022.
- [203] A. More and S. More. Improved statistic to identify strongly lensed gravitational wave events. *Mon. Not. Roy. Astron. Soc.*, 515(1):1044–1051, June 2022.
- [204] C. Baraldo, A. Hosoya, and T. T. Nakamura. Gravitationally induced interference of gravitational waves by a rotating massive object. *Phys. Rev. D*, 59:083001, April 1999.
- [205] A. K. Meena, A. Mishra, A. More, et al. Gravitational Lensing of Gravitational Waves: Probability of Microlensing in Galaxy-Scale Lens Population. *arXiv e-prints*, page arXiv:2205.05409, May 2022.
- [206] P. C. Peters. Index of refraction for scalar, electromagnetic, and gravitational waves in weak gravitational fields. *Phys. Rev. D*, 9:2207–2218, April 1974.
- [207] H. J. Witt. Investigation of high amplification events in light curves of gravitationally lensed quasars. *Astron. Astrophys.*, 236:311, September 1990.
- [208] R. Kormann, P. Schneider, and M. Bartelmann. Isothermal elliptical gravitational lens models. *Astron. Astrophys.*, 284:285–299, March 1994.
- [209] L. V. E. Koopmans, A. Bolton, T. Treu, et al. The Structure and Dynamics of Massive Early-Type Galaxies: On Homology, Isothermality, and Isotropy Inside One Effective Radius. *Astrophys. J. Lett.*, 703(1):L51–L54, September 2009.
- [210] S. Birrer and A. Amara. Lenstronomy: multi-purpose gravitational lens modelling software package. *arXiv e-prints*, page 1803.09746, March 2018.
- [211] S. Birrer, A. J. Shajib, D. Gilman, et al. lenstronomy ii: A gravitational lensing software ecosystem. *Journal of Open Source Software*, 6(62):3283, June 2021.
- [212] A. K. Mehta, C. K. Mishra, V. Varma, and P. Ajith. Accurate inspiral-merger-ringdown gravitational waveforms for nonspinning black-hole binaries including the effect of subdominant modes. *Phys. Rev. D*, 96(12):124010, December 2017.

[213] S. V. Dhurandhar and M. Tinto. Astronomical observations with a network of detectors of gravitational waves. I - Mathematical framework and solution of the five detector problem. *Mon. Not. Roy. Astron. Soc.*, 234:663–676, October 1988.

[214] R. K. L. Lo and I. Magaña Hernandez. A Bayesian statistical framework for identifying strongly-lensed gravitational-wave signals. *arXiv e-prints*, page 2104.09339, April 2021.

[215] X. Liu, I. Magaña Hernandez, and J. Creighton. Identifying strong gravitational-wave lensing during the second observing run of Advanced LIGO and Advanced Virgo. *Astrophys. J.*, 908(1):97, February 2021.

[216] M. Bojarski, D. Del Testa, D. Dworakowski, et al. End to End Learning for Self-Driving Cars. *arXiv e-prints*, page arXiv:1604.07316, April 2016.

[217] T. Raviteja and S.I. Rajay Vedaraj. Global Image Segmentation Process using Machine Learning algorithm & Convolution Neural Network method for Self- Driving Vehicles. *arXiv e-prints*, page arXiv:2010.13294, October 2020.

[218] I. U. Haq. An overview of deep learning in medical imaging. *arXiv e-prints*, page arXiv:2202.08546, February 2022.

[219] E. Dadzie and K. Kwakye. Developing a Machine Learning Algorithm-Based Classification Models for the Detection of High-Energy Gamma Particles. *arXiv e-prints*, page arXiv:2111.09496, November 2021.

[220] Y. M. Seo, P. F. Goldsmith, V. Tolls, et al. Applications of Machine Learning Algorithms in Processing Terahertz Spectroscopic Data. *Journal of Astronomical Instrumentation*, 9:2050011, January 2020.

[221] T. Miener, D. Nieto, R. López-Coto, et al. The performance of the MAGIC telescopes using deep convolutional neural networks with CTLearn. November 2022.

[222] A. Vaswani, N. Shazeer, N. Parmar, et al. Attention Is All You Need. *arXiv e-prints*, page arXiv:1706.03762, June 2017.

[223] OpenAI. Chatgpt: Optimizing language models for dialogue. <https://openai.com/blog/chatgpt/>, 2022.

[224] N. I. Galanis, P. Vafiadis, K. G. Mirzaev, and G. A. Papakostas. Machine Learning Meets Natural Language Processing – The story so far. *arXiv e-prints*, page arXiv:2104.10213, March 2021.

[225] A. Torfi, R. A. Shirvani, Y. Keneshloo, et al. Natural Language Processing Advancements By Deep Learning: A Survey. *arXiv e-prints*, page arXiv:2003.01200, March 2020.

[226] F. Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386–408, November 1958.

[227] D. Hendrycks and K. Gimpel. Gaussian Error Linear Units (GELUs). *arXiv e-prints*, page arXiv:1606.08415, June 2016.

[228] P.T. de Boer and D.P. Kroese. Estimating buffer overflows in three stages using cross-entropy. In *Proceedings of the Winter Simulation Conference*, volume 1, pages 301–309 vol.1, January 2002.

- [229] D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. *arXiv e-prints*, page arXiv:1412.6980, December 2014.
- [230] I. Loshchilov and F. Hutter. Decoupled Weight Decay Regularization. *arXiv e-prints*, page arXiv:1711.05101, November 2017.
- [231] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv e-prints*, page arXiv:1409.1556, September 2014.
- [232] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, May 2017.
- [233] J. Redmon, S. Divvala, R. Girshick, et al. You Only Look Once: Unified, Real-Time Object Detection. *arXiv e-prints*, page arXiv:1506.02640, June 2015.
- [234] K. He, X. Zhang, S. Ren, et al. Deep Residual Learning for Image Recognition. *arXiv e-prints*, page arXiv:1512.03385, December 2015.
- [235] A. Bochkovskiy, C. Wang, and H. M. Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv e-prints*, page arXiv:2004.10934, April 2020.
- [236] C. Szegedy, S. Ioffe, V. Vanhoucke, et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *arXiv e-prints*, page arXiv:1602.07261, February 2016.
- [237] S. Ioffe and C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv e-prints*, page arXiv:1502.03167, February 2015.
- [238] G. Papamakarios, E. Nalisnick, D. Jimenez Rezende, et al. Normalizing Flows for Probabilistic Modeling and Inference. *arXiv e-prints*, page arXiv:1912.02762, December 2019.
- [239] C. Huang, D. Krueger, A. Lacoste, et al. Neural Autoregressive Flows. *arXiv e-prints*, page arXiv:1804.00779, April 2018.
- [240] L. Dinh, D. Krueger, and Y. Bengio. NICE: Non-linear Independent Components Estimation. *arXiv e-prints*, page arXiv:1410.8516, October 2014.
- [241] L. Dinh, J. Sohl-Dickstein, and S. Bengio. Density estimation using Real NVP. *arXiv e-prints*, page arXiv:1605.08803, May 2016.
- [242] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, et al. Neural Ordinary Differential Equations. *arXiv e-prints*, page arXiv:1806.07366, June 2018.
- [243] M.F. Hutchinson. A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics - Simulation and Computation*, 19(2):433–450, May 1990.
- [244] R. P. Adams, J. Pennington, M. J. Johnson, et al. Estimating the Spectral Density of Large Implicit Matrices. *arXiv e-prints*, page arXiv:1802.03451, February 2018.
- [245] W. Grathwohl, R. T. Q. Chen, J. Bettencourt, et al. FFJORD: Free-form Continuous Dynamics for Scalable Reversible Generative Models. *arXiv e-prints*, page arXiv:1810.01367, October 2018.
- [246] U. Mutze. An asynchronous leapfrog method II. *arXiv e-prints*, page arXiv:1311.6602, November 2013.

[247] J. Zhuang, N. C. Dvornek, S. Tatikonda, et al. MALLI: A memory efficient and reverse accurate integrator for Neural ODEs. *arXiv e-prints*, page arXiv:2102.04668, February 2021.

[248] C. Winkler, D. Worrall, E. Hoogeboom, et al. Learning Likelihoods with Conditional Normalizing Flows. *arXiv e-prints*, page arXiv:1912.00042, November 2019.

[249] T. Charnock, G. Lavaux, and B. D. Wandelt. Automatic physical inference with information maximizing neural networks. *Phys. Rev. D*, 97(8):083004, April 2018.

[250] N. Halko, P. Martinsson, and J. A. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *arXiv e-prints*, page arXiv:0909.4061, September 2009.

[251] K. Cannon, A. Chapman, C. Hanna, et al. Singular value decomposition applied to compact binary coalescence gravitational-wave signals. *Phys. Rev. D*, 82:044025, August 2010.

[252] C. M. Bishop and N. M. Nasrabadi. Pattern recognition and machine learning. *J. Electronic Imaging*, 16:049901, August 2006.

[253] P. T. H. Pang, O. A. Hannuksela, T. Dietrich, et al. Lensed or not lensed: Determining lensing magnifications for binary neutron star mergers from a single detection. *Mon. Not. Roy. Astron. Soc.*, 495(4):3740–3750, July 2020.

[254] A. K. Y. Li, R. K. L. Lo, S. Sachdev, et al. Finding diamonds in the rough: Targeted Sub-threshold Search for Strongly-lensed Gravitational-wave Events. *arXiv e-prints*, page 1904.06020, April 2019.

[255] C. McIsaac, D. Keitel, T. Collett, et al. Search for strongly lensed counterpart images of binary black hole mergers in the first two LIGO observing runs. *Phys. Rev. D*, 102(8):084031, October 2020.

[256] J. Januart, K. Haris, and O.A.H Hannuksela. GOLUM: a software for rapid strongly-lensed gravitational wave parameter estimation . <https://github.com/lemnisi2/golum>, 2022.

[257] I. Mandel, W. M. Farr, and J. R. Gair. Extracting distribution parameters from multiple uncertain observations with selection biases. *Mon. Not. Roy. Astron. Soc.*, 486(1):1086–1093, June 2019.

[258] T. E. Collett and D. J. Bacon. Compound lensing: Einstein Zig-Zags and high multiplicity lensed images. *Mon. Not. Roy. Astron. Soc.*, 456(2):2210–2220, December 2016.

[259] H. Dahle, M. D. Gladders, K. Sharon, et al. SDSS J2222+2745: A Gravitationally Lensed Sextuple Quasar with a Maximum Image Separation of 15.”1 Discovered in the Sloan Giant Arcs Survey. *Astrophys. J.*, 773(2):146, August 2013.

[260] T. E. Collett, E. Buckley-Geer, H. Lin, et al. Core or Cusps: The Central Dark Matter Profile of a Strong Lensing Cluster with a Bright Central Image at Redshift 1. *Astrophys. J.*, 843(2):148, July 2017.

[261] C. García-Quirós, M. Colleoni, S. Husa, et al. Multimode frequency-domain model for the gravitational wave signal from nonprecessing black-hole binaries. *Phys. Rev. D*, 102(6):064002, September 2020.

[262] J. Calderón Bustillo, S. Husa, A. M. Sintes, et al. Impact of gravitational radiation higher order modes on single aligned-spin gravitational wave searches for binary black holes. *Phys. Rev. D*, 93(8):084019, April 2016.

- [263] V. Varma and P. Ajith. Effects of nonquadrupole modes in the detection and parameter estimation of black hole binaries with nonprecessing spins. *Phys. Rev. D*, 96(12):124024, December 2017.
- [264] S. Mukherjee, T. Broadhurst, J. M. Diego, et al. Impact of astrophysical binary coalescence time-scales on the rate of lensed gravitational wave events. *Mon. Not. Roy. Astron. Soc.*, 506(3):3751–3759, July 2021.
- [265] F. Xu, J. M. Ezquiaga, and D. E. Holz. Please Repeat: Strong Lensing of Gravitational Waves as a Probe of Compact Binary and Galaxy Populations. *Astrophys. J.*, 929(1):9, April 2022.
- [266] LIGO Scientific Collaboration and Virgo Collaboration and KAGRA Collaboration. The population of merging compact binaries inferred using gravitational waves through GWTC-3 - Data release (Version v1). <https://zenodo.org/record/5655785>, November 2021.
- [267] LIGO Scientific Collaboration, VIRGO Collaboration, and KAGRA Collaboration. GWTC-2.1: Deep Extended Catalog of Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run. *arXiv e-prints*, page 2108.01045, August 2021.
- [268] LIGO Scientific Collaboration and Virgo Collaboration and KAGRA Collaboration. GWTC-2.1: Deep Extended Catalog of Compact Binary Coalescences Observed by LIGO and Virgo During the First Half of the Third Observing Run - Parameter Estimation Data Release (Version v2). <https://zenodo.org/record/5117703>, July 2021.
- [269] LIGO Scientific Collaboration and Virgo Collaboration and KAGRA Collaboration. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run — Parameter estimation data release (Version v1). <https://zenodo.org/record/5546663>, November 2021.
- [270] H. Jeffreys. *The theory of probability*. OUP Oxford, August 1998.
- [271] L. Dai, T. Venumadhav, and K. Sigurdson. Effect of lensing magnification on the apparent distribution of black hole mergers. *Phys. Rev. D*, 95(4):044011, February 2017.
- [272] R. Barkana. Fast calculation of a family of elliptical mass gravitational lens models. *Astrophys. J.*, 502:531, August 1998.
- [273] S. M. C. Yeung, M. H. Y. Cheung, J. A. J. Gais, et al. Microlensing of type II gravitational-wave macroimages. December 2021.
- [274] C. Meegan, G. Lichti, P. Bhat, et al. The fermi gamma-ray burst monitor. *Astrophys. J.*, 702(1):791, August 2009.
- [275] V. Savchenko, C. Ferrigno, E. Kuulkers, et al. INTEGRAL detection of the first prompt gamma-ray signal coincident with the gravitational-wave event GW170817. *Astrophys. J. Lett.*, 848(2):L15, October 2017.
- [276] P. Cowperthwaite, E. Berger, V. Villar, et al. The electromagnetic counterpart of the binary neutron star merger LIGO/Virgo GW170817. II. UV, optical, and near-infrared light curves and comparison to kilonova models. *Astrophys. J. Lett.*, 848(2):L17, October 2017.
- [277] M. Soares-Santos, A. Palmese, W. Hartley, et al. First measurement of the hubble constant from a dark standard siren using the dark energy survey galaxies and the ligo/virgo binary-black-hole merger gw170814. *Astrophys. J. Lett.*, 876(1):L7, April 2019.

[278] M. Fishbach, R. Gray, I. Hernandez Magaña, et al. A standard siren measurement of the hubble constant from gw170817 without the electromagnetic counterpart. *Astrophys. J. Lett.*, 871(1):L13, January 2019.

[279] E. Berti, K. Yagi, and N. Yunes. Extreme gravity tests with gravitational waves from compact binary coalescences:(i) inspiral–merger. *General Relativity and Gravitation*, 50(4):46, April 2018.

[280] LIGO Scientific Collaboration and Virgo Collaboration. Tests of general relativity with the binary black hole signals from the ligo-virgo catalog gwtc-1. *Phys. Rev. D*, 100(10):104036, November 2019.

[281] LIGO Scientific Collaboration and Virgo Collaboration. A gravitational-wave standard siren measurement of the hubble constant. *Nature*, 551(7678):85–8, October 2017.

[282] LIGO Scientific Collaboration and Virgo Collaboration. Low-latency gravitational-wave alerts for multimessenger astronomy during the second advanced ligo and virgo observing run. *Astrophys. J.*, 875(2):161, April 2019.

[283] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. Search for Gravitational Waves Associated with Gamma-Ray Bursts Detected by Fermi and Swift during the LIGO–Virgo Run O3b. *Astrophys. J.*, 928(2):186, April 2022.

[284] M. Saleem, A. Pai, K. Misra, et al. Rates of short-GRB afterglows in association with binary neutron star mergers. *Mon. Not. Roy. Astron. Soc.*, 475(1):699–707, March 2018.

[285] B. Patricelli, M. G. Bernardini, M. Mapelli, et al. Prospects for multi-messenger detection of binary neutron star mergers in the fourth LIGO–Virgo–KAGRA observing run. *Mon. Not. Roy. Astron. Soc.*, 513(3):4159–4168, July 2022. [Erratum: Mon. Not. Roy. Astron. Soc. 514, 3395 (2022)].

[286] S. Ronchini, M. Branchesi, G. Oganesyan, et al. Perspectives for multimessenger astronomy with the next generation of gravitational-wave detectors and high-energy satellites. *Astron. Astrophys.*, 665:A97, September 2022.

[287] S. Sachdev, R. Magee, C. Hanna, et al. An Early-warning System for Electromagnetic Follow-up of Gravitational-wave Events. *Astrophys. J. Lett.*, 905(2):L25, December 2020.

[288] R. Magee, D. Chatterjee, L. P. Singer, et al. First Demonstration of Early Warning Gravitational-wave Alerts. *Astrophys. J. Lett.*, 910(2):L21, April 2021.

[289] G. Baltus, J. Janquart, M. Lopez, et al. Convolutional neural networks for the detection of the early inspiral of a gravitational-wave signal. *Phys. Rev. D*, 103:102003, May 2021.

[290] H. Yu, R. X. Adhikari, R. Magee, et al. Early warning of coalescing neutron-star and neutron-star-black-hole binaries from the nonstationary noise background using neural networks. *Phys. Rev. D*, 104(6):062004, September 2021.

[291] G. Baltus, J. Janquart, M. Lopez, et al. Convolutional neural network for gravitational-wave early alert: Going down in frequency. *Phys. Rev. D*, 106(4):042002, May 2022.

[292] L. P. Singer and L. R. Price. Rapid Bayesian position reconstruction for gravitational-wave transients. *Phys. Rev. D*, 93(2):024013, January 2016.

- [293] A. Delaunoy, A. Wehenkel, T. Hinderer, et al. Lightning-Fast Gravitational Wave Parameter Inference through Neural Amortization. *arXiv e-prints*, page 2010.12931, October 2020.
- [294] S. R. Green, C. Simpson, and J. Gair. Gravitational-wave parameter estimation with autoregressive neural network flows. *Phys. Rev. D*, 102(10):104057, November 2020.
- [295] S. R. Green and J. Gair. Complete parameter inference for GW150914 using deep learning. *Mach. Learn. Sci. Tech.*, 2(3):03LT01, June 2021.
- [296] M. Dax, S. R. Green, J. Gair, et al. Group equivariant neural posterior estimation. *arXiv e-prints*, page 2111.13139, November 2021.
- [297] M. Dax, S. R. Green, J. Gair, et al. Real-Time Gravitational Wave Science with Neural Posterior Estimation. *Phys. Rev. Lett.*, 127(24):241103, December 2021.
- [298] M. Dax, S. R. Green, J. Gair, et al. Neural Importance Sampling for Rapid and Reliable Gravitational-Wave Inference. *arXiv e-prints*, page 2210.05686, October 2022.
- [299] M. J. Williams, J. Veitch, and C. Messenger. Nested sampling with normalizing flows for gravitational-wave inference. *Phys. Rev. D*, 103(10):103006, May 2021.
- [300] A. Kolmus, G. Baltus, J. Janquart, et al. Fast sky localization of gravitational waves using deep learning seeded importance sampling. *Phys. Rev. D*, 106(2):023032, July 2022.
- [301] D. N. Burrows, J. Hill, J. Nousek, et al. The Swift X-ray Telescope. *Space Sci. Rev.*, 120:165, August 2005.
- [302] M. Nicholl, E. Berger, D. Kasen, et al. The Electromagnetic Counterpart of the Binary Neutron Star Merger LIGO/Virgo GW170817. III. Optical and UV Spectra of a Blue Kilonova from Fast Polar Ejecta. *Astrophys. J. Lett.*, 848(2):L18, October 2017.
- [303] B. D. Metzger. Welcome to the Multi-Messenger Era! Lessons from a Neutron Star Merger and the Landscape Ahead. *arXiv e-prints*, page 1710.05931, October 2017.
- [304] B. D. Metzger and A. L. Piro. Optical and X-ray emission from stable millisecond magnetars formed from the merger of binary neutron stars. *Mon. Not. Roy. Astron. Soc.*, 439(4):3916–3930, March 2014.
- [305] R. Ciolfi and D. M. Siegel. Short gamma-ray bursts in the “time-reversal” scenario. *Astrophys. J. Lett.*, 798(2):L36, December 2014.
- [306] D. M. Siegel and R. Ciolfi. Electromagnetic Emission from Long-lived Binary Neutron Star Merger Remnants. II. Lightcurves and Spectra. *Astrophys. J. Lett.*, 819(1):15, March 2016.
- [307] E. R. Most and A. A. Philippov. Electromagnetic Precursors to Gravitational-wave Events: Numerical Simulations of Flaring in Pre-merger Binary Neutron Star Magnetospheres. *Astrophys. J. Lett.*, 893(1):L6, April 2020.
- [308] Tomonori Totani. Cosmological Fast Radio Bursts from Binary Neutron Star Mergers. *Publications of the Astronomical Society of Japan*, 65(5), October 2013. L12.
- [309] J.-S. Wang, Y.-P. Yang, X.-F. Wu, et al. Fast Radio Bursts from the Inspiral of Double Neutron Stars. *Astrophys. J. Lett.*, 822(1):L7, May 2016.

[310] V. I. Dokuchaev and Yu. N. Eroshenko. Recurrent fast radio bursts from collisions of neutron stars in the evolved stellar clusters. *arXiv e-prints*, page 1701.02492, January 2017.

[311] D. George and E.A. Huerta. Deep Neural Networks to Enable Real-time Multimessenger Astrophysics. *Phys. Rev. D*, 97(4):044039, February 2018.

[312] D. George and E.A. Huerta. Deep Learning for Real-time Gravitational Wave Detection and Parameter Estimation: Results with Advanced LIGO Data. *Phys. Lett. B*, 778:64–70, March 2018.

[313] H. Gabbard, M. Williams, F. Hayes, et al. Matching matched filtering with deep networks for gravitational-wave astronomy. *Phys. Rev. Lett.*, 120(14):141103, April 2018.

[314] W. Wei and E. A. Huerta. Deep learning for gravitational wave forecasting of neutron star mergers. *Phys. Lett. B*, 816:136185, May 2021.

[315] P. G. Krastev. Real-time detection of gravitational waves from binary neutron stars using artificial neural networks. *Phys. Lett. B*, 803:135330, April 2020.

[316] T. Gebhard, N. Kilbertus, G. Parascandolo, et al. Convwave: Searching for gravitational waves with fully convolutional neural nets. In *Workshop on Deep Learning for Physical Sciences (DLPS) at the 31st Conference on Neural Information Processing Systems (NIPS)*, pages 1–6, December 2017.

[317] S. Chatterji, L. Blackburn, G. Martin, et al. Multiresolution techniques for the detection of gravitational-wave bursts. *Class. Quant. Grav.*, 21:S1809–S1818, September 2004.

[318] C. Messick, K. Blackburn, P. Brady, et al. Analysis framework for the prompt discovery of compact binary mergers in gravitational-wave data. *Phys. Rev. D*, 95(4):042001, February 2017.

[319] A. Samajdar, J. Januart, C. Van Den Broeck, et al. Biases in parameter estimation from overlapping gravitational-wave signals in the third-generation detector era. *Phys. Rev. D*, 104(4):044003, August 2021.

[320] C. Cutler and E. E. Flanagan. Gravitational waves from merging compact binaries: How accurately can one extract the binary’s parameters from the inspiral wave form? *Phys. Rev. D*, 49:2658–2697, March 1994.

[321] C.M. Biwer, C. D. Capano, S. De, et al. PyCBC Inference: A Python-based parameter estimation toolkit for compact binary coalescence signals. *Publ. Astron. Soc. Pac.*, 131(996):024503, January 2019.

[322] A. Buonanno, Y. Chen, and M. Vallisneri. Detecting gravitational waves from precessing binaries of spinning compact objects: Adiabatic limit. *Phys. Rev. D*, 67:104025, May 2003. [Erratum: *Phys. Rev. D*, 74, 029904 (2006)].

[323] A. Paszke, S. Gross, F. Massa, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., December 2019.

[324] T. Fawcett. An introduction to ROC Analysis. *Pattern Recognition Lett.*, 27:861–874, June 2006.

[325] S. Sachdev, S. Caudill, H. Fong, et al. The GstLAL Search Analysis Methods for Compact Binary Mergers in Advanced LIGO’s Second and Advanced Virgo’s First Observing Runs. *arXiv e-prints*, page 1901.08580, January 2019.

[326] LIGO Scientific Collaboration, Virgo Collaboration, and KARGRA Collaboration. Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog. *Astrophys. J. Lett.*, 913(1):L7, May 2021.

[327] LIGO Scientific Collaboration, Virgo Collaboration, and KAGRA Collaboration. Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO, Advanced Virgo and KAGRA. *Living Rev. Rel.*, 21(1):3, September 2018.

[328] M. López, I. Di Palma, M. Drago, et al. Deep learning for core-collapse supernova detection. *Phys. Rev. D*, 103:063011, March 2021.

[329] Y. Bengio, J. Louradour, R. Collobert, et al. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48, June 2009.

[330] P. Petrov, L.P. Singer, M.W. Coughlin, et al. Data-driven expectations for electromagnetic counterpart searches based on ligo/virgo public alerts. *Astrophys. J.*, 924(2):54, January 2022.

[331] D. M. Macleod, J. Areeda, S.B. Coughlin, et al. GWpy: A Python package for gravitational-wave astrophysics. *SoftwareX*, 13:100657, January 2021.

[332] B. Kiziltan, A. Kottas, S.E. Thorsett, et al. The neutron star mass distribution. *Astrophys. J.*, 778(1):66, November 2013.

[333] D. De Boer, P. T. and Kroese, S. Mannor, et al. A tutorial on the cross-entropy method. *Annals of operations research*, 134(1):19–67, February 2005.

[334] A. H. Nitz, M. Schafer, and T. Dal Canton. Gravitational-wave merger forecasting: Scenarios for the early detection and localization of compact-binary mergers with ground-based observatories. *Astrophys. J. Lett.*, 902(2):L29, October 2020.

[335] Z. Doctor, R. Kessler, K. Herner, et al. A search for optical emission from binary black hole merger gw170814 with the dark energy camera. *Astrophys. J. Lett.*, 873(2):L24, March 2019.

[336] R. Perna, D. Lazzati, and W. Farr. Limits on electromagnetic counterparts of gravitational-wave-detected binary black hole mergers. *Astrophys. J.*, 875(1):49, January 2019.

[337] L. Barack, V. Cardoso, S. Nissanke, et al. Black holes, gravitational waves and fundamental physics: a roadmap. *Class. Quant. Grav.*, 36(14):143001, June 2019.

[338] K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, August 2012.

[339] X. Fan, J. Li, X. Li, et al. Applying deep neural networks to the detection and space parameter estimation of compact binary coalescence with a network of gravitational wave detectors. *Science China Physics, Mechanics & Astronomy*, 62(6):1–8, February 2019.

[340] N. I. Fisher, T. Lewis, and B. J. J. Embleton. *Statistical analysis of spherical data*. Cambridge university press, August 1993.

[341] R. L. Russell and C. Reale. Multivariate uncertainty in deep learning. *IEEE Transactions on Neural Networks and Learning Systems*, October 2021.

[342] I. Loshchilov and F. Hutter. SGDR: Stochastic Gradient Descent with Warm Restarts. In *ICLR: International Conference on Learning Representations*, March 2017.

[343] T. D. Gebhard, N. Kilbertus, I. Harry, et al. Convolutional neural networks: A magic bullet for gravitational-wave detection? *Phys. Rev. D*, 100(6):063015, September 2019.

[344] C. Chatterjee, L. Wen, K. Vinsen, et al. Using deep learning to localize gravitational wave sources. *Phys. Rev. D*, 100(10):103025, November 2019.

[345] A. J. K. Chua and M. Vallisneri. Learning bayesian posteriors with neural networks for gravitational-wave inference. *Phys. Rev. Lett.*, 124(4):041102, January 2020.

[346] P. Ajith, M. Hewitson, and I. S. Heng. Null-stream veto for two co-located detectors: Implementation issues. *Class. Quant. Grav.*, 23:S741–S749, September 2006.

[347] I. C. F. Wong, P. T. H. Pang, R. K. L. Lo, et al. Null-stream-based Bayesian Unmodeled Framework to Probe Generic Gravitational-wave Polarizations. *arXiv e-prints*, page 2105.09485, May 2021.

[348] B. Goncharov, A. H. Nitz, and J. Harms. Utilizing the null stream of the Einstein Telescope. *Phys. Rev. D*, 105(12):122007, June 2022.

[349] S. Hild, S. Chelkowski, and A. Freise. Pushing towards the ET sensitivity using 'conventional' technology. *arXiv e-prints*, page 0810.0604, November 2008.

[350] S. Köppel, L. Bovard, and L. Rezzolla. A General-relativistic Determination of the Threshold Mass to Prompt Collapse in Binary Neutron Star Mergers. *Astrophys. J. Lett.*, 872(1):L16, February 2019.

[351] A. Bauswein, T. W. Baumgarte, and H. T. Janka. Prompt merger collapse and the maximum mass of neutron stars. *Phys. Rev. Lett.*, 111(13):131101, September 2013.

[352] A. Puecher, T. Dietrich, K. W. Tsang, et al. Unraveling information about supranuclear-dense matter from the complete binary neutron star coalescence process using future gravitational-wave detector networks. *arXiv e-prints*, page 2210.09259, October 2022.

[353] S. Morisaki. Accelerating parameter estimation of gravitational waves from compact binary coalescence using adaptive frequency resolutions. *Phys. Rev. D*, 104(4):044062, August 2021.

[354] B. Zackay, L. Dai, and T. Venumadhav. Relative Binning and Fast Likelihood Evaluation for Gravitational Wave Parameter Estimation. *arXiv e-prints*, page 1806.08792, June 2018.

[355] L. Dai, T. Venumadhav, and B. Zackay. Parameter Estimation for GW170817 using Relative Binning. *arXiv e-prints*, page 1806.08793, June 2018.

[356] N. Leslie, L. Dai, and G. Pratten. Mode-by-mode relative binning: Fast likelihood estimation for gravitational waveforms with spin-orbit precession and multiple harmonics. *Phys. Rev. D*, 104(12):123030, December 2021.

[357] J. Januart. RelativeBilbying: a package for relative binning with bilby. <https://github.com/lemnis12/relativebilbying>, 2022.

[358] S. Gao, F. Hayes, S. Croke, et al. Quantum algorithm for gravitational-wave matched filtering. *Phys. Rev. Res.*, 4(2):023006, April 2022.

[359] T. Regimbau and S. A. Hughes. Gravitational-wave confusion background from cosmological compact binaries: Implications for future terrestrial detectors. *Phys. Rev. D*, 79:062002, March 2009.

- [360] E. Pizzati, S. Sachdev, A. Gupta, et al. Toward inference of overlapping gravitational-wave signals. *Phys. Rev. D*, 105(10):104016, May 2022.
- [361] P. Relton and V. Raymond. Parameter estimation bias from overlapping binary black hole events in second generation interferometers. *Phys. Rev. D*, 104(8):084039, October 2021.
- [362] A. Antonelli, O. Burke, and J. R. Gair. Noisy neighbours: inference biases from overlapping gravitational-wave signals. *Mon. Not. Roy. Astron. Soc.*, 507(4):5069–5086, August 2021.
- [363] Y. Himemoto, A. Nishizawa, and A. Taruya. Impacts of overlapping gravitational-wave signals on the parameter estimation: Toward the search for cosmological backgrounds. *Phys. Rev. D*, 104(4):044010, August 2021.
- [364] B. Sathyaprakash, M. Abernathy, F. Acernese, et al. Scientific Objectives of Einstein Telescope. *Class. Quant. Grav.*, 29:124013, June 2012. [Erratum: *Class. Quant. Grav.* 30, 079501 (2013)].
- [365] B. S. Sathyaprakash, B. F. Schutz, and C. Van Den Broeck. Cosmography with the Einstein Telescope. *Class. Quant. Grav.*, 27:215006, September 2010.
- [366] C. Van Den Broeck. Compact binary coalescence and the science case for Einstein Telescope. In *12th Marcel Grossmann Meeting on General Relativity*, pages 1682–1685, March 2010.
- [367] W. Zhao, C. Van Den Broeck, D. Baskaran, et al. Determination of Dark Energy by the Einstein Telescope: Comparing with CMB, BAO and SNIa Observations. *Phys. Rev. D*, 83:023005, January 2011.
- [368] M. Punturo, M. Abernathy, F. Acernese, et al. The third generation of gravitational wave observatories and their science reach. *Class. Quant. Grav.*, 27:084007, April 2010.
- [369] M. Sathyaprakash, B. abd Abernathy, F. Acernese, et al. Scientific Potential of Einstein Telescope. In *46th Rencontres de Moriond on Gravitational Waves and Experimental Gravity*, pages 127–136, August 2011.
- [370] C. Van Den Broeck. Astrophysics, cosmology, and fundamental physics with compact binary coalescence and the Einstein Telescope. *J. Phys. Conf. Ser.*, 484:012008, March 2014.
- [371] B. S. Sathyaprakash, A. Buonanno, L. Lehner, et al. Extreme Gravity and Fundamental Physics. *Bulletin of the AAS*, (3), May 2019.
- [372] S. Vitale, W. M. Farr, K. K. Y. Ng, et al. Measuring the star formation rate with gravitational waves from binary black holes. *Astrophys. J.*, 886(1):L1, November 2019.
- [373] K. K. Y. Ng, S. Vitale, W. M. Farr, et al. Probing multiple populations of compact binaries with third-generation gravitational-wave detectors, May 2021.
- [374] C. Cutler and J. Harms. Big bang observer and the neutron-star-binary subtraction problem. *Phys. Rev. D*, 73(4), February 2006.
- [375] J. Harms, C. Mahrdt, M. Otto, et al. Subtraction-noise projection in gravitational-wave detector networks. *Phys. Rev. D*, 77(12), June 2008.
- [376] T. Regimbau, M. Evans, N. Christensen, et al. Digging deeper: Observing primordial gravitational waves below the binary-black-hole-produced stochastic background. *Phys. Rev. Lett.*, 118(15), April 2017.

[377] S. Sachdev, T. Regimbau, and B.S. Sathyaprakash. Subtracting compact binary foreground sources to reveal primordial gravitational-wave backgrounds. *Phys. Rev. D*, 102(2), July 2020.

[378] A. Sharma and J. Harms. Searching for cosmological gravitational-wave backgrounds with third-generation detectors in the presence of an astrophysical foreground. *Phys. Rev. D*, 102(6), September 2020.

[379] S. Biscoveanu, C. Talbot, E. Thrane, et al. Measuring the primordial gravitational-wave background in the presence of astrophysical foregrounds. *Phys. Rev. Lett.*, 125(24), December 2020.

[380] P. Relton, A. Virtuoso, S. Bini, et al. Addressing the challenges of detecting time-overlapping compact binary coalescences. *Phys. Rev. D*, 106(10):104045, November 2022.

[381] S. Vitale. Three observational differences for binary black holes detections with second- and third-generation gravitational-wave detectors. *Phys. Rev. D*, 94(12), December 2016.

[382] S. Vitale and M. Evans. Parameter estimation for binary black holes with networks of third-generation gravitational-wave detectors. *Physical Review D*, 95(6), March 2017.

[383] S. Vitale and C. Whittle. Characterization of binary black holes by heterogeneous gravitational-wave networks. *Phys. Rev. D*, 98(2), July 2018.

[384] LIGO Scientific Collaboration and VIRGO Collaboration. GW151226: Observation of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence. *Phys. Rev. Lett.*, 116(24):241103, June 2016.

[385] B. Hoang, S. Naoz, and K. Kremer. Neutron Star–Black Hole Mergers from Gravitational-wave Captures. *Astrophys. J.*, 903(1):8, November 2020.

[386] T. Regimbau, T. Dent, W. Del Pozzo, et al. A Mock Data Challenge for the Einstein Gravitational-Wave Telescope. *Phys. Rev. D*, 86:122001, December 2012.

[387] M. Safarzadeh, E. Berger, K. K.-Y. Ng, et al. Measuring the delay time distribution of binary neutron stars. II. Using the redshift distribution from third-generation gravitational wave detectors network. *Astrophys. J. Lett.*, 878(1):L13, June 2019.

[388] K. Belczynski, T. Ryu, R. Perna, et al. On the likelihood of detecting gravitational waves from Population III compact object binaries. *Mon. Not. Roy. Astron. Soc.*, 471(4):4702–4721, July 2017.

[389] M. Oguri. Effect of gravitational lensing on the distribution of gravitational waves from distant binary black hole mergers. *Mon. Not. Roy. Astron. Soc.*, 480(3):3842–3855, August 2018.

[390] Planck Collaboration. Planck 2013 results. XVI. Cosmological parameters. *Astron. Astrophys.*, 571:A16, November 2014.

[391] T. P. Robitaille, E. J. Tollerud, P. Greenfield, et al. Astropy: A community python package for astronomy. *Astron. Astrophys.*, 558:A33, September 2013.

[392] A. M. Price-Whelan, B. M. Sipocz, H. M. Günther, et al. The astropy project: Building an open-science project and status of the v2.0 core package. *The Astronomical Journal*, 156(3):123, August 2018.

- [393] S. Kastha, M. Saleem, and K.G. Arun. Imprints of the redshift evolution of double neutron star merger rate on the signal to noise ratio distribution. *Mon. Not. Roy. Astron. Soc.*, 496(1):523–531, July 2020.
- [394] A. Freise, S. Chelkowski, S. Hild, et al. Triple Michelson Interferometer for a Third-Generation Gravitational Wave Detector. *Class. Quant. Grav.*, 26:085012, April 2009.
- [395] S. Hild, S. Chelkowski, A. Freise, et al. A Xylophone Configuration for a third Generation Gravitational Wave Detector. *Class. Quant. Grav.*, 27:015003, December 2009.
- [396] B. S. Sathyaprakash, E. Belgacem, D. Bertacca, et al. Cosmology and the early universe, May 2019.
- [397] R. X. Adhikari, K. Arai, A. F. Brooks, et al. A cryogenic silicon interferometer for gravitational-wave detection. *Class. Quant. Grav.*, 37(16):165003, July 2020.
- [398] E. E. Flanagan and T. Hinderer. Constraining neutron star tidal Love numbers with gravitational wave detectors. *Phys. Rev. D*, 77:021502, January 2008.
- [399] J. Vines, E. E. Flanagan, and T. Hinderer. Post-1-newtonian tidal effects in the gravitational waveform from binary inspirals. *Phys. Rev. D*, 83:084051, April 2011.
- [400] T. Damour, A. Nagar, and L. Villain. Measurability of the tidal polarizability of neutron stars in late-inspiral gravitational-wave signals. *Phys. Rev. D*, 85:123007, June 2012.
- [401] W. Del Pozzo, T. G. F. Li, M. Agathos, et al. Demonstrating the feasibility of probing the neutron star equation of state with second-generation gravitational wave detectors. *Phys. Rev. Lett.*, 111(7):071101, August 2013.
- [402] M. Agathos, J. Meidam, W. Del Pozzo, et al. Constraining the neutron star equation of state with gravitational wave signals from coalescing binary neutron stars. *Phys. Rev. D*, 92(2):023012, July 2015.
- [403] L. Wade, J. D. E. Creighton, E. Ochsner, et al. Systematic and statistical errors in a bayesian approach to the estimation of the neutron-star equation of state using advanced gravitational wave detectors. *Phys. Rev. D*, 89(10):103012, May 2014.
- [404] T. Dietrich, A. Samajdar, S. Khan, et al. Improving the NRTidal model for binary neutron star systems. *Phys. Rev. D*, 100(4):044003, August 2019.
- [405] LIGO Scientific Collaboration. LIGO Algorithm Library - LALSuite. free software (GPL), 2018.
- [406] C. Van Den Broeck and A. S. Sengupta. Binary black hole spectroscopy. *Class. Quant. Grav.*, 24:1089–1114, March 2007.
- [407] A.N. Kolmogorov. Sulla Determinazione Empirica di Una Legge di Distribuzione. *G. Inst. Ital. Attuari*, 4:83, December 1933.
- [408] N. Smirnov. Table for Estimating the Goodness of Fit of Empirical Distributions. *Ann. Math. Statist.*, 19:279–281, June 1948.
- [409] A. Akmal, V. R. Pandharipande, and D. G. Ravenhall. The Equation of state of nucleon matter and neutron star structure. *Phys. Rev. C*, 58:1804–1828, September 1998.

[410] LIGO Scientific Collaboration and Virgo Collaboration. Properties of the binary neutron star merger GW170817. *Phys. Rev. X*, 9(1):011001, January 2019.

[411] A. Samajdar and T. Dietrich. Waveform systematics for binary neutron star gravitational wave signals: effects of the point-particle baseline and tidal descriptions. *Phys. Rev. D*, 98(12):124030, December 2018.

[412] A. Samajdar and T. Dietrich. Waveform systematics for binary neutron star gravitational wave signals: Effects of spin, precession, and the observation of electromagnetic counterparts. *Phys. Rev. D*, 100(2):024046, July 2019.

[413] J. Crowder and N. J. Cornish. LISA source confusion. *Phys. Rev. D*, 70:082004, October 2004.

[414] N. J. Cornish and J. Crowder. LISA data analysis using Markov chain Monte Carlo methods. *Phys. Rev. D*, 72:043005, August 2005.

[415] N. J. Cornish and E. K. Porter. The Search for supermassive black hole binaries with LISA. *Class. Quant. Grav.*, 24:5729–5755, December 2007.

[416] J. Crowder and N. Cornish. A Solution to the Galactic Foreground Problem for LISA. *Phys. Rev. D*, 75:043008, February 2007.

[417] T. B. Littenberg. Detection pipeline for galactic binaries in LISA data. *Phys. Rev. D*, 84:063009, September 2011.

[418] T. Robson and N. Cornish. Impact of galactic foreground characterization on a global analysis for the LISA gravitational wave observatory. *Class. Quant. Grav.*, 34(24):244002, November 2017.

[419] T. Littenberg, N. Cornish, K. Lackeos, et al. Global Analysis of the Gravitational Wave Signal from Galactic Binaries. *Phys. Rev. D*, 101(12):123021, June 2020.

[420] A. Petiteau, S. Babak, A. Sesana, et al. Resolving multiple supermassive black hole binaries with pulsar timing arrays II: genetic algorithm implementation. *Phys. Rev. D*, 87(6):064036, March 2013.

[421] S. Wu and A. H. Nitz. A mock data study for 3G ground-based detectors: the performance loss of matched filtering due to correlated confusion noise. *arXiv e-prints*, page 2209.03135, September 2022.

[422] Q. Hu and J. Veitch. Accumulating errors in tests of general relativity with the Einstein Telescope: overlapping signals and inaccurate waveforms. *arXiv e-prints*, page 2210.04769, October 2022.

[423] B. Zhou, L. Reali, E. Berti, et al. Subtracting Compact Binary Foregrounds to Search for Subdominant Gravitational-Wave Backgrounds in Next-Generation Ground-Based Observatories. *arXiv e-prints*, page 2209.01310, September 2022.

[424] B. Zhou, L. Reali, E. Berti, et al. Compact Binary Foreground Subtraction in Next-Generation Ground-Based Observatories. *arXiv e-prints*, page 2209.01221, September 2022.

[425] L. Reali, A. Antonelli, R. Cotesta, et al. The impact of confusion noise on golden binary neutron-star events in next-generation terrestrial observatories. *arXiv e-prints*, page 2209.13452, September 2022.

- [426] J. Janquart, T. Baka, A. Samajdar, et al. Parameter estimation methods for analyzing overlapping gravitational wave signals in the third-generation detector era. *arXiv e-prints*, page arXiv:2211.01304, November 2022.
- [427] B. P. Abbott, R. Abbott, T. D. Abbott, et al. Exploring the sensitivity of next generation gravitational wave detectors. *Class. Quant. Grav.*, 34(4):044001, January 2017.
- [428] J. Langendorff, A. Kolmus, J. Janquart, et al. Normalizing flows as an avenue to study overlapping gravitational wave signals. *arXiv e-prints*, page arXiv:2211.15097, November 2022.
- [429] D. Jimenez Rezende and S. Mohamed. Variational Inference with Normalizing Flows. *arXiv e-prints*, page arXiv:1505.05770, May 2015.
- [430] D. P. Kingma, T. Salimans, R. Jozefowicz, et al. Improving Variational Inference with Inverse Autoregressive Flow. *arXiv e-prints*, page arXiv:1606.04934, June 2016.
- [431] G. Papamakarios, T. Pavlakou, and I. Murray. Masked Autoregressive Flow for Density Estimation. *arXiv e-prints*, page arXiv:1705.07057, May 2017.
- [432] K. Cranmer, J. Brehmer, and G. Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, April 2020.
- [433] G. Papamakarios and I. Murray. Fast ε -free Inference of Simulation Models with Bayesian Conditional Density Estimation. *arXiv e-prints*, page arXiv:1605.06376, May 2016.
- [434] D. Ha, A. M. Dai, and Q. V. Le. Hypernetworks. In *International Conference on Learning Representations*, September 2017.